

Emergent Phenomena in Correlated Matter

Eva Pavarini, Erik Koch, and Ulrich Schollwöck (Eds.)

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

German Research School for
Simulation Sciences GmbH

**Lecture Notes of the Autumn School
Correlated Electrons 2013**

Eva Pavarini, Erik Koch, and Ulrich Schollwöck (Eds.)

Emergent Phenomena in Correlated Matter

Autumn School organized by
the Forschungszentrum Jülich
and the German Research School
for Simulation Sciences

at Forschungszentrum Jülich
23 – 27 September 2013

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation

Band / Volume 3

ISSN 2192-8525

ISBN 978-3-89336-884-6

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data are available in the
Internet at <http://dnb.d-nb.de>.

Publisher: Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Druckerei Schloemer, Düren

Copyright: Forschungszentrum Jülich 2013

Distributor: Forschungszentrum Jülich
Zentralbibliothek, Verlag
D-52425 Jülich
Phone +49 (0)2461 61-5368 · Fax +49 (0)2461 61-6103
e-mail: zb-publikation@fz-juelich.de
Internet: <http://www.fz-juelich.de>

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 3

ISSN 2192-8525
ISBN 978-3-89336-884-6

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JUWEL) at
<http://www.fz-juelich.de/zb/juwel>

Neither this book nor any part of it may be reproduced or transmitted in any form or by any
means, electronic or mechanical, including photocopying, microfilming, and recording, or by any
information storage and retrieval system, without permission in writing from the publisher.

Contents

Preface

Introduction

1. Density Functional Theory for Emergents
Robert O. Jones
2. Many-Electron States
Erik Koch
3. Magnetism: Models and Mechanisms
Eva Pavarini
4. The Variational Cluster Approximation
Robert Eder
5. Magnetism: From Stoner to Hubbard
Alexander I. Lichtenstein
6. Monte Carlo Methods with Applications to Spin Systems
Werner Krauth
7. Monte Carlo Simulations of Quantum Spin Models
Stefan Wessel
8. Quantum Theory of Molecular Magnetism
Jürgen Schnack
9. Recent Advances in Experimental Research on High-Temperature Superconductivity
Bernhard Keimer
10. Strongly Correlated Superconductivity
André-Marie S. Tremblay
11. Superconductivity: 2D Physics, Unknown Mechanisms, Current Puzzles
Warren E. Pickett
12. Density Functional Perturbation Theory and Electron Phonon Coupling
Rolf Heid
13. Eliashberg Theory
Giovanni A.C. Ummarino
14. Path Integral Methods for Continuum Quantum Systems
David M. Ceperley
15. Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems
Shiwei Zhang
16. DMRG: Ground States, Time Evolution, and Spectral Functions
Ulrich Schollwöck
17. Entanglement and Tensor Network States
Jens Eisert

Index

Preface

Emergent phenomena are the hallmark of many-body systems, and yet to unravel their nature remains one of the central challenges in condensed-matter physics. In order to advance our understanding it is crucial to learn from the different manifestations of emergence as well as from the interplay of different emergent phases, such as magnetism and superconductivity. For addressing such problems, it is necessary to master a broad spectrum of techniques from traditionally separate branches of research, ranging from *ab-initio* approaches based on density-functional theory to advanced many-body methods, electron-lattice coupling and dynamics. In these lecture notes we analyze emergence in some of its major manifestations in the solid-state and compare methodologies used to address specific aspects. The aim of the school is to introduce advanced graduate students and up to the essence of emergence and to the modern approaches for modeling emergent properties of correlated matter.

A school of this size and scope requires support and help from many sources. We are very grateful for all the financial and practical support we have received. The Institute for Advanced Simulation and the German Research School for Simulation Sciences at the Forschungszentrum Jülich provided the funding and were vital for the organization of the school and the production of this book. The Institute for Complex Adaptive Matter (ICAM) offered travel support for international speakers and participants.

The nature of a school makes it desirable to have the lecture-notes available already during the lectures. In this way the participants get the chance to work through the lectures thoroughly while they are given. We are therefore extremely grateful to the lecturers that, despite a tight schedule, provided their manuscripts in time for the production of this book. We are confident that the lecture notes collected here will not only serve the participants of the school but will also be useful for other students entering the exciting field of strongly correlated materials.

We thank Mrs. H. Lexis of the Forschungszentrum Jülich Verlag as well as Mr. D. Laufenberg and Mrs. C. Reisen of the Graphische Betriebe for providing their expert support in producing the present volume on a tight schedule. We heartily thank our students and postdocs that helped in proofreading the manuscripts, often on quite short notice: Michael Baumgärtel, Khaldoon Ghanem, Evgeny Gorelov, Esmaeel Sarvestani, Amin Kiani Sheikhabadi, Joaquin Miranda, German Ulm, Guoren Zhang, and in particular Hunter Sims.

Finally, our special thanks go to Dipl.-Ing. R. Hölzle for his invaluable advice on the innumerable questions concerning the organization of such an endeavour and to Mrs. L. Snyders and Mrs. E. George for expertly handling all practical issues.

Eva Pavarini, Erik Koch, and Ulrich Schollwöck

July 2013

I Introduction

Eva Pavarini and Erik Koch
Institute for Advanced Simulation
German Research School for Simulation Sciences
Forschungszentrum Jülich

Contents

1	Emergent phenomena	2
2	Paradigmatic cases	8
2.1	Antiferromagnetism	9
2.2	Superconductivity	11
3	Overview of the school	13

1 Emergent phenomena

The concept of emergence arose in arguments about the biological basis of consciousness that were sparked by Darwin's theory of evolution. It was formalized by the philosopher George Henry Lewes in his 1875 *Problems of Life and Mind* [1]. But John Stuart Mill, around whom a successful school of *British Emergentism* developed, had already noted in his 1843 *System of Logic* [2]

All organized bodies are composed of parts, similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but the phenomena of life, which result from the juxtaposition of those parts in a certain manner, bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents. To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself.

While it was hotly debated whether the constitutive principles of biology are reducible to those of physics and chemistry, the importance of emergence in physics was recognized only much later. From the inception of their science, physicists have been fascinated most with finding the ultimate laws of the universe. There is no doubt that our present understanding of the world owes a lot to this *reductionist* approach. That one and the same force, gravitation, explains how an apple falls and how the planets move around the sun is an astonishing discovery. Other ideas such as the existence of atoms, a finite number of particles from which all normal matter is made, revolutionized our worldview. The understanding of the differences between atoms, which led to the periodic table, opened the path to quantum mechanics and gave chemistry a microscopic basis. At the same time, the landscape grew more complex. It became clear that atoms are not at all the fundamental particles, but that they are composed of yet more elementary particles, electrons, neutrons, and protons. It was understood that a new force holds the nuclei together, which was later still understood to be the consequence of a more fundamental force, the strong force, between more elementary particles, the quarks. The stellar successes of the reductionist approach led a large part of the scientific community to think that the new grand challenge was to find the theory of everything, in which the ultimate elementary particles are identified and where all known forces are unified. From this point of view, everything will ultimately be explained in the same framework. Among the many successes of this approach, the most striking and perhaps least celebrated revelation is, however, that *every fundamental theory hides a more fundamental one* and that the target seems to elude us, today's elementary particles and fundamental forces becoming tomorrow's bound states and effective interactions. So the question is whether an ultimate fundamental theory of everything would solve all problems. A glimpse into the history of condensed-matter physics hints at the answer: not really. At the theory layer of chemistry and solid-state phenomena, the fundamental particles and interac-

tions have actually been known since the first quarter of the 20th century. Most of solid-state physics and chemistry can indeed be described by the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} |\Psi\rangle = H|\Psi\rangle \quad (1)$$

where the Hamiltonian for a set of atomic nuclei $\{\alpha\}$ with atomic numbers $\{Z_\alpha\}$ and masses $\{M_\alpha\}$ and their accompanying electrons $\{i\}$ is given, in atomic units, by

$$H = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_\alpha \frac{1}{2M_\alpha} \nabla_\alpha^2 - \sum_{\alpha,i} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \frac{1}{2} \sum_{ii'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} + \frac{1}{2} \sum_{\alpha\alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|}.$$

As Laughlin and Pines pointed out [3], here we have the theory of (almost) everything (ToaE). Why then are we still working on condensed-matter physics? The history of 20th century condensed-matter physics is full of experimental discoveries showing genuinely novel behaviors that were not only unanticipated by theory, but took decades to fully clarify, even though, in principle, the equations, the ToaE, were well known. Examples are magnetism, superconductivity, the Kondo effect, or the quantum-Hall effect in its integer and fractional forms. Some of the mysteries discovered in the last century are still not or not fully clarified, such as high-temperature superconductivity and heavy-fermion behavior; new ones will certainly come to light in the years to come. The difficulties are not merely technical or computational, but fundamental. As P.W. Anderson pointed out in the now very famous article *More is Different* [4], when many particles interact, even if we know the type of interaction, and even if the interaction is as simple as a two-body term, totally surprising results can emerge: gold is made of gold atoms; but a single atom of gold is neither a metal nor does it appear in a golden color. Likewise, an Fe atom is not ferromagnetic, nor does a Pb atom superconduct at low temperatures. Perhaps it is easier to grasp the concept of emergence by going back to its origin. As Anderson writes [5]

The idea of emergence began in biology, where it was realized that the course of evolution is one long story of novel properties emerging out of the synergy of large collections of simpler systems: cells from biomolecules, organisms from cells, societies from organisms. But that emergence could occur in physics was a novel idea. Perhaps not totally novel: I heard the great evolutionist Ernst Mayr claiming that 30 or 40 years ago, when he described emergence to Niels Bohr, Bohr said; “but we have that in physics as well! – physics is all emergent”, but at the time, as usual, only Bohr knew what he meant.

Thus the challenge in condensed-matter physics is to understand the *behavior of complex systems*, starting from the apparently *simple underlying theory*, the Schrödinger equation. This is what Anderson defines the *Complex Way*, in contrast to the *Glamor Way* travelled by high-energy physics in the search of the ultimate theory. Complexity is tamed by *universality*: it is rather the consequence of many-body correlations – one particle influenced by all the others – than of the specific type of interaction. New complex entities can then form under well defined

conditions, such as high pressure, low temperature, or in the presence of magnetic fields, often in radically different systems. These entities are stable in a certain regime, in which they represent the fundamental “particles” – but actually are effective- or quasi-particles. They feel effective interactions among themselves, yielding on a higher level yet other states of matter. Can we predict such states without experimental facts, just from the equations? As Anderson writes [5]

The structure of Nature is so completely hierarchical, built up from emergence upon emergence, that the very concepts and categories on which the next stage can take place are themselves arbitrary.

Remarkably, modern molecular biology has made enormous progress by identifying genes as fundamental entities. Still, the evolutionary biologist Ernst Mayr rejects the reductionist approach that evolutionary pressure acts on single genes, arguing that it instead acts on organisms, that genes influence each other, and that accounting for this influence is essential. Even if we take genes as the fundamental entities, their definition might appear to the outsider as airy as that of quasiparticles in physics or what we before called complex entities. It is clear that, whatever they are, they are composite rather than fundamental objects; but it is also clear that they are a better starting point than their more fundamental components. It would, e.g., not be of much help for molecular biology to start from the Schrödinger equation. Returning to the layer of electrons and nuclei, the message is that at a given energy scale new intermediate layers can form at which novel emergent behavior occurs. But, when dealing with systems made of $N \sim 10^{23}$ particles, to predict novel emergent behavior is hardly possible. Again, quoting Anderson [5]

How can you predict the result when you can't predict what you will be measuring?

The classical path to discoveries of new states of matter is, with few precious exceptions, led by experiment. It is the recognition of a paradox, an experimental result apparently contradicting our well established theories, which leads to the identification of new phenomena; Nature has to provide us some strong hint. And it is hardly the *ab-initio* path, from the ToaE to the real world, that leads us to understanding the physics of the new phenomenon. It is rather the identification of mechanisms, which often is based on simple models, apparently wild ideas, and a good measure of approximation.

It is natural to ask ourselves how emergent phenomena arise. The formation of layers of physical theory is best understood in terms of the idea of *renormalization*. The exact solution of the Schrödinger equation involves phenomena at many energy scales. However, at low energy – meaning the energy window on which we are focusing – high-energy states play a small role. They can only be reached via virtual excitations, which have a time-scale proportional to $\hbar/\Delta E$, where ΔE is the excitation energy. Thus it is not necessary to account for them in full detail. It is more meaningful to downfold them and work in the subspace in which only low-energy states and, if sizable, the low-energy effects of these virtual excitations are taken into account. Between different branches of physics, this separation is quite clear-cut: in condensed-matter

physics, we do not need quarks to describe atoms, not even neutrons and protons. Instead we can simply talk about their bound states, the nuclei as immutable objects. In these cases, the high-energy states are so far away that the only effect of the downfolding is the emergent object, e.g., the atomic nucleus. For intermediate energy scales the scenario can become much more complicated; the effect of the downfolding is to generate *effective Hamiltonians*, the new effective theory in that energy window. The effective Hamiltonian is typically made of the original Hamiltonian, however restricted to the low-energy Hilbert space and with renormalized parameters, plus new interactions; in the smaller Hilbert space some degrees of freedom are frozen, and the effective Hamiltonian typically can be rewritten in term of new entities, stable in that subspace; examples are quasiparticles, Cooper pairs, or local spins. The major difficulty in condensed-matter is that there is always a chance of crossing the boundaries between effective theories, e.g., in a phase transition to a new state of matter. Still it is possible to identify truly emergent regimes from their robustness or universality. Often the cooperative behavior of a many-particle system is surprisingly independent of the details of their realization – their substrate. The Kondo effect, for example, was initially found in diluted magnetic alloys and ascribed to the antiferromagnetic exchange interaction of the localized impurity-spins with the spin-density of the conduction electrons at the impurity site. But it has been realized in a number of systems in which there are no local magnetic moments, such as in quantum dots or in carbon nanotubes. Recently it has even been shown to be intimately related to the metal-insulator transition of the Mott type, a connection that is at the core of the dynamical mean-field theory approach (DMFT) [6, 7]. Another striking example is conventional superconductivity, explained via the BCS theory. First observed in simple metals at very low temperatures, the same kind of phenomena has been found in liquid ^3He , which becomes superfluid below a certain critical temperature and even in systems as exotic as neutron stars [8]. Returning to less massive systems, the mechanism of the metal-insulator transition in transition-metal oxides is typically described via the Hubbard model, which in recent years has also been employed to model the behavior of very different systems: ultra-cold atoms in optical lattices.

But how do we go from the Schrödinger equation to emergent properties? It is certainly tempting to start from the exact many-body wave function, as we know a straightforward prescription for calculating observables from it. But for what system should we calculate that exact wave function? We are certainly not interested in the properties of just one particular sample with its unique arrangement of atoms. What we are interested in are the properties of a *material*, i.e., a class of samples that can be reproducibly manufactured. Any single one of these samples will be quite different from the others in terms of its microscopic details. So, even if we could calculate them, the exact ground state wave functions for two such samples would, for all practical purposes, be orthogonal. Chipping off an atom from a bulk of gold does not change its characteristics. Thus, we are not really interested in the exact solution for some particular situation, but in general properties. Emergent properties abstract from the idiosyncrasies of a particular realization. A typical idealization is the *thermodynamic limit*, where we assume that the number of particles $N \rightarrow \infty$, even though any real sample can only be made of a finite number of atoms. This lets us exploit the advantages of the continuum over discrete sets. Only

in the thermodynamic limit can we define a phase transition where every actual sample shows a mere crossover. Only in an infinite system can we speak of a continuum of excited states where for every actual sample there is only a finite number of excitations with a distribution that depends critically on its size. Another important idealization is the perfect-crystal approximation, in which we assume that defects are of minor importance for the physical properties of interest and where, in particular, we abstract from surface effects by introducing periodic boundary conditions. It is of course crucial to keep in mind that these idealizations only make sense for properties that can be transferred from the idealization to the real materials. Experimentally we can, for all practical purposes, distinguish metals and insulator and observe phase transitions in crystals. Thus the thermodynamic limit is a good starting point. Conversely, trying to describe a bulk system with a finite number of atoms, we eventually get lost in irrelevant details. The importance of getting rid of irrelevant details has been succinctly expressed by Lipkin [9]

On the other hand, the exact solution of a many-body problem is really irrelevant since it includes a large mass of information about the system which although measurable in principle is never measured in practice. In this respect, the quantum-mechanical many-body problem resembles the corresponding problem in classical statistical mechanics. Although it is possible in principle to solve Newton's equations of motion and obtain the classical trajectories of all the particles in the system for all times, only a small part of this information is relevant to anything that is measurable in practice. Classical statistical mechanics uses a statistical description in which measurable macroscopic quantities such as temperature, pressure and entropy play an important role. An incomplete description of the system is considered to be sufficient if these measurable quantities and their behavior are described correctly.

Thus, approximate methods that grasp the essential details are bound to be more successful than exact methods – if available [10] – since they shield us from all the irrelevant information. So it is not the “Great Solid State Physics Dream Machine” [5] that we should be after, or in the words of Wigner and Seitz [11]

If one had a great calculating machine, one might apply it to the problem of solving the Schrödinger equation for each metal and obtain thereby the interesting physical quantities, such as the cohesive energy, the lattice constant, and similar parameters. It is not clear, however, that a great deal would be gained by this. Presumably the results would agree with the experimentally determined quantities and nothing vastly new would be learned from the calculation. It would be preferable instead to have a vivid picture of the behavior of the wave functions, a simple description of the essence of the factors which determine cohesion and an understanding of the origins of variation in properties [...].

Lipkin concludes that it is actually misleading to think that our job is to find approximations to the exact solution of the Schrödinger equation [9]

In fact, many treatments of the quantum-mechanical many-body problem give the misleading impression that they are indeed methods for obtaining approximations to the exact solution of the Schrödinger equation.

A better approach is to develop approximate schemes for calculating, for idealized systems, reduced quantities that do not provide complete information of the system but still allow us to calculate experimentally accessible quantities, such as Green functions or response functions. This route is successfully taken by density-functional theory (DFT) [12], which uses the many-body electron density as a variable, or by the dynamical mean-field theory, which is based instead on Green functions. The great success of DFT is that, via the Kohn-Sham equation, the problem of finding the electron density of the original many-electron system is reduced to that of calculating it for an auxiliary one-electron problem, whose Hamiltonian has to be determined self-consistently. The electrons in the auxiliary problem feel a one-electron potential whose strength and shape is determined not only by the nuclei which define the lattice but also by all the other electrons – which is why self-consistency is needed. Of course, even though DFT is in principle an exact ground-state theory, we know only approximate forms of the DFT potential, such as the local-density approximation (LDA) and its extension. Nevertheless, if the approximate form is good enough, we can perform *ab-initio* calculations, i.e., calculate many properties of a system specified only by the atomic positions and the type of atoms. In an emergent world, it would, however, be very surprising if the LDA always worked. The reason is a fundamental one. If objects qualitatively different from the quasi-electrons, on which the LDA is built, can form, this approach is bound to fail even qualitatively. This is what happens in strongly correlated materials. Remarkably, however, the LDA is so successful that DFT can be considered the *standard model* of solid state physics that is used to understand and even predict the properties of whole classes of materials. Strongly correlated systems are not only characterized by the fact that the LDA fails to describe them. More importantly, their properties are very sensitive to small changes in external fields or doping, and hence they are characterized by surprisingly large effects, such as colossal magneto-resistance, high-temperature superconductivity, and the like. This suggests that a variety of different layers can easily form in which new fundamental entities exist and interact. Hence, for strongly correlated systems, it is particularly unlikely that a single approximation can be sufficient for explaining all phenomena.

One exemplary failure of the LDA is the Mott metal-insulator transition. Within the LDA, metals have partially filled bands while insulators are characterized by a band gap. The latter can also arise because of long-range magnetic order. The same concept of metal and insulator remains in place if we use approaches in which many-body effects are taken into account on a static mean-field level such as the Hartree-Fock (HF) method. Thus the existence of materials with partially filled bands that are paramagnetic insulators is a paradox in the context of LDA or HF. It can, however, be understood using a simple model, the *Hubbard model*

$$H = -t \sum_{\sigma} \sum_{\langle ii' \rangle} c_{i\sigma}^{\dagger} c_{i'\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (2)$$

In this model, the metal-insulator transition at half-filling is a consequence of a large t/U ratio, i.e., an on-site Coulomb repulsion which is large with respect to the band-width, determined by the hopping t . Although the mechanism behind the paramagnetic insulating phase had been proposed about 60 years ago by Nevil Mott, it is only recently, through the dynamical mean-

field theory [6, 7], that we can indeed describe that physics quantitatively. DMFT yields an approximate solution of the Hubbard model in which spatial correlations are neglected and only dynamical effects are taken into account. The DMFT solution of the Hubbard model shows that with increasing t/U the quasiparticle masses increase and eventually diverge; the Mott transition is associated with the corresponding divergence of the real part of the self energy.

It is remarkable that a model as simple as the Hubbard model, bare of all the complications or details of the real material, can be of any relevance for systems as different as NiO, LaTiO₃ and KCuF₃. LDA-based studies on weakly correlated materials suggest that this cannot possibly be the case. When the LDA works, we typically need the full details of the electronic structure to explain the structure of the Fermi surfaces, the lattice, or the chemical bond. In fact, model Hamiltonians might grasp the substance of a phenomenon, such as the nature of the Mott transition, but they are not sufficient to account for the varieties of its manifestations, and eventually will fail in explaining new paradoxes that are found when this variety is explored. The Mott mechanism can explain the existence of Mott insulators, but it does not tell us why SrVO₃ is metallic while the very similar YTiO₃ and LaTiO₃ are insulators, or why the gap is much larger in YTiO₃ than in LaTiO₃, although they have a similar LDA band width. Even if we know that Mott physics is the right starting point, we have to augment the simple Hubbard model to describe reality. But how? If we could solve it, the original Hamiltonian contains all details, but, as we have discussed, they are too many, and thus they tell us nothing. The crucial point is to disentangle the important features from all the irrelevant details. This is the true challenge in condensed-matter physics. In the case of SrVO₃, YTiO₃ and LaTiO₃, it turns out that structural distortions and a tiny crystal field splitting of t_{2g} levels play the crucial role [13]. Not surprisingly, there is no systematic way of determining which details do matter for explaining a certain behavior in a given system. This process relies on our intuition about the mechanisms, and it brings the work of a physicist rather close to that of an artist, requiring proper taste and judgement. The good news is that for Mott-like system the DMFT method turns out to be a very flexible approach. It has been combined with DFT in the LDA+DMFT approach [14], whose steady development in the last ten years allows us to solve more and more realistic Hubbard-like models and thus to test ideas and approximation in a realistic context. And it is not difficult to imagine that in the next 20 years LDA+DMFT codes will probably become as flexible and versatile as modern DFT codes. We have to keep in mind that, although this constitutes impressive progress, as in the case of the LDA, it is very unlikely that a single approximation will solve all the paradoxes. New ones will certainly be found, and will require us to extend the theory, to think differently, to go away from the well known path, to look for new mechanisms. The end of physics is unlikely to come any time soon.

2 Paradigmatic cases

At the focus of this year's school are two paradigmatic examples of emergent phenomena, antiferromagnetism and electron-phonon driven superconductivity. We briefly reconstruct the main steps that led to the unraveling of their mystery and discuss their emergent aspects.

2.1 Antiferromagnetism

At the beginning of the 20th century, magnetism was the subject of intense debate. The theoretical scenario was dominated by the ingenious concept of Weiss' molecular field theory of ferromagnetism [15]. Of course magnetic moments in matter could only be understood quantum mechanically and it was Heisenberg, who proposed that the ferromagnetic coupling between magnetic moments is due to the *Coulomb exchange* between electrons [16]. The latter yields an interaction of the form

$$H = \frac{\Gamma}{2} \sum_{ii'} \mathbf{S}_i \cdot \mathbf{S}_{i'}, \quad (3)$$

with $\Gamma < 0$. Néel [17] extended Weiss' theory to the case of a site-dependent molecular field and found the antiferromagnetic state as the mean-field solution of the $\Gamma > 0$ Heisenberg model below a critical temperature, T_N . Antiferromagnetism is one of the precious exceptions to the rule that condensed-matter physics is essentially led by experiment: The experimental proof of the existence of antiferromagnetism came only much later, in 1951, when Shull and Smart measured via elastic neutron scattering sharp new Bragg peaks below a critical temperature in the transition-metal oxide MnO [18]. Even in the case in which they are actually predicted, however, emergent phenomena are rarely as simple as in the original theoretical proposal. The paradox at the time was that the exact solution of the antiferromagnetic one-dimensional Heisenberg chain, obtained by Bethe, yields a ground state with total-spin zero, a condition not satisfied by the Néel antiferromagnetic state. Later on, this paradox was solved by the observation of Anderson that, in a perfect antiferromagnet, quantum fluctuations would restore the symmetry, but in a real system weak perturbations, defects, or an external magnetic field can suppress them; these quantum fluctuations however imply the existence of cooperative excitations, *spin waves*. This is a consequence of the Goldstone theorem, which states that soft bosonic excitations have to be present whenever a continuous symmetry is broken. Antiferromagnetism turns out to be, indeed, a representative example of a so-called *broken symmetry* state, a state in which the electrons choose not to have the same symmetry of the Hamiltonian that govern their behavior, in this specific case the continuous spin-rotation symmetry of the Heisenberg model Eq. (3). Remarkably, some of the ideas developed in the context of broken symmetry, such as *Goldstone bosons*, were taken over by high-energy physics [5], and have driven the search for the Higgs boson [19]. It is worth pointing out another emergent aspect associated with the Heisenberg model that brings us back to the basics of quantum mechanics. Where do the local spins \mathbf{S}_i come from? There are apparently no such local spins in the original Schrödinger Hamiltonian, the ToaE of solid-state physics. The existence of local spins becomes immediately clear, however, if we consider an idealized atom described by the Hamiltonian

$$H = U n_{\uparrow} n_{\downarrow}.$$

This Hamiltonian has four eigenstates, which can be labeled as $|N, S, S_z\rangle$, where N is the total number of electrons and S_z , the z component of the total spin,

$$\begin{aligned} |0, 0, 0\rangle &= |0\rangle \\ |1, \frac{1}{2}, \uparrow\rangle &= c_{\uparrow}^{\dagger} |0\rangle \\ |1, \frac{1}{2}, \downarrow\rangle &= c_{\downarrow}^{\dagger} |0\rangle \\ |2, 0, 0\rangle &= c_{\uparrow}^{\dagger} c_{\downarrow}^{\dagger} |0\rangle \end{aligned}$$

These states result from the electron-electron interaction and the Pauli principle. In this simple example the energy of the atomic states depends only on the total number of electrons; thus we label it with $E(N)$, with $E(0) = 0$, $E(1) = 0$, and $E(2) = U$. If the atom is full or empty, its total spin is zero; if instead the idealized atom is occupied by one electron, it behaves as a local $S = 1/2$. Let us consider now a half-filled system described by the one-band Hubbard model, Eq. (2), in which a set of idealized atoms of the type just discussed form a lattice. In this model the electrons can, in principle, hop from site to site, gaining kinetic energy $\propto -t$; each hopping process will, however, cost the Coulomb energy

$$E(2) + E(0) - 2E(1) = U.$$

If the ratio between kinetic energy gain and Coulomb energy loss, $\propto t/U$, is small enough double-occupations are unlikely and each site is filled on average with ~ 1 electron. Then spins remain stable in the crystal, and the overall effect of the virtual excitations to $N = 0$ and $N = 2$ states is an effective exchange interaction between the spins. We can calculate the effective exchange coupling by downfolding the $N = 0$ and $N = 2$ high-energy states; if we follow this procedure [20] we find an effective antiferromagnetic Heisenberg interaction with

$$\Gamma = -\frac{4t^2}{U}.$$

This is an example of the *kinetic exchange* mechanism. It plays an important role in the physics of transition-metal oxides. We are now in the position to discuss emergence at work. The first lesson is that spins are by themselves emergent objects. They are the result of the interplay of the Pauli principle and Coulomb repulsion. Furthermore, within the kinetic exchange mechanism, they interact because of virtual excitations to high-energy states, in this case those with $N = 2$ and $N = 0$. Then, below a certain temperature T_N , because of the interactions between these emergent entities, a new cooperative emergent state, the antiferromagnetic Néel state, arises; an example of emergence built on emergence. Finally, spins only “live” on certain energy scales. At energy scales comparable with t/U excitations to empty and doubly occupied states (charge fluctuations) become likely and it is no longer possible to describe the system by a simple Heisenberg model; this happens, for example, when we want to study the Hubbard bands in photoemission spectra. Thus, increasing the energy or the temperature we cross the boundary to a different layer and change the effective theory. In the higher layer of theory the spins are the fundamental particles and the Heisenberg model becomes an effective theory of everything. In the lower layer we have to account for the charge degrees of freedom, and the effective theory of everything is the Hubbard model.

2.2 Superconductivity

The discovery of metals with infinite conductivity [21] by Kamerlingh Onnes in 1911 came as a genuine surprise. It took almost 60 years to find an explanation, years in which brilliant minds tried hard to solve the riddle and yet failed. This failure by itself is a strong indication that superconductivity is an emergent phenomenon. Many experimental facts were added along the way; we just mention two of the most significant. The first was the Meissner effect in 1933 [22], the spontaneous expulsion of a magnetic field, somewhat similar to perfect diamagnetism. The most crucial observation was perhaps the discovery of the isotope effect in 1950 [23]. From the theory side, the decisive development was the concept of electron pairs as developed by Leon Cooper in late 1956 [24]. Cooper realized that, in the presence of an arbitrarily weak electron-electron attraction, $-V$, two electrons with energy just above the Fermi surface of a metal will form a bound-state that could destabilize the Fermi surface itself. Cooper's pair creation operator is defined as

$$b_{\text{CP}}^\dagger = \sum_{\mathbf{k}} \lambda_{\mathbf{k}} c_{\mathbf{k}\downarrow}^\dagger c_{-\mathbf{k}\uparrow}^\dagger.$$

Since a Cooper pair is, to first approximation, a boson, Cooper pairs can in principle all occupy the same state, as it happens in Bose-Einstein condensation. Based on these ideas, Bardeen, Cooper and Schrieffer elaborated the theory of superconductivity. They identified the superconducting state as a coherent state, the eigenstate of Cooper's pair annihilation operator b_{CP} . In Fock space such state can be easily written in product form

$$|\Psi_{\text{BCS}}\rangle = e^{b_{\text{CP}}^\dagger} |0\rangle = \prod_{\mathbf{k}} \left(1 + \lambda_{\mathbf{k}} c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger \right) |0\rangle,$$

where $|0\rangle$ is the electron vacuum. The microscopic mechanism that leads to the pairing in conventional superconductors is the electron-phonon coupling; for electrons right above the Fermi surface the resulting electron-electron coupling is attractive. The BCS Hamiltonian has then the form

$$H_{\text{BCS}} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\mathbf{k}'} V_{\mathbf{k},\mathbf{k}'} c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\downarrow} c_{-\mathbf{k}'\uparrow},$$

where $\varepsilon_{\mathbf{k}}$ is the dispersion of the electrons, and $V_{\mathbf{k},\mathbf{k}'}$ the electron-electron interaction. In the simplest version of the BCS theory one can assume that the coupling $V_{\mathbf{k},\mathbf{k}'}$ is isotropic. Thus we can make the following approximation

$$V_{\mathbf{k},\mathbf{k}'} \sim \begin{cases} -\bar{V} & \text{for } |\varepsilon_{\mathbf{k}}|, |\varepsilon_{\mathbf{k}'}| < \varepsilon_D \\ 0 & \text{otherwise} \end{cases}$$

where ε_D defines the small energy window in which the potential is attractive. The superconducting gap $\Delta(T)$ is then given by the solution of the BCS gap equation; at $T = 0$

$$\Delta(0) = 2\varepsilon_D e^{-1/\rho_{\sigma}(\varepsilon_F)\bar{V}}, \quad (4)$$

where $\rho_\sigma(\varepsilon_F)$ is the density of states per spin at the Fermi level, and $\rho_\uparrow(\varepsilon_F) = \rho_\downarrow(\varepsilon_F)$. There is a universal relation between the gap and the critical temperature,

$$\frac{2\Delta(0)}{T_c} \approx 3.528.$$

In superconductivity the continuous symmetry that is broken is the gauge symmetry related to the conservation of charge; in the broken symmetry state the *phase* of the wavefunction is the new physical parameter. The Cooper pairs are emergent objects, which interact to form a macroscopic condensate: the superconducting state.

Nowadays, we call superconductors that can be explained within the BCS theory and its simple extensions *conventional superconductors*. This does, however, not mean that they do not hold surprises. The discovery of MgB_2 in 2001 [25], with a T_c as high as 40 K, was totally unanticipated, in particular in such a comparatively simple binary compound. It immediately sparked an intense search for similar conventional materials with perhaps even higher T_c . Remarkably, MgB_2 is not an exotic material; at the time of the discovery it was available to many laboratories. In principle we did, once more, have the theory – we could have predicted it. But again, nobody thought in advance that MgB_2 could be such a remarkable system. And even if we had used our theoretical tools, would we have predicted that MgB_2 is a high- T_c conventional superconductor? Probably not: it turns out that MgB_2 is less simple than one might think. To understand it we have to account for multiple bands and gap anisotropies, typically neglected in the standard version of the theory of conventional superconductors. Thus this is a case in which details that are usually negligible play an essential role.

Another, totally different surprise had arrived earlier, in 1986, with the discovery of superconductivity with $T_c = 40$ K in La_2CuO_4 . The finding was so unexpected that the title of the paper [26] that won Bednorz and Müller a Nobel prize conveys the author's doubts: *Possible high T_c superconductivity in the Ba-La-Cu-O system*. In a relatively short time, an entire family of CuO_2 -layered superconducting materials was identified, the high-temperature superconducting cuprates (HTSCs). Within the HTSC family, the maximum value of T_c rose rapidly to ~ 130 K. It quickly became clear that these new materials differ substantially from conventional superconductors and the mechanism for high-temperature superconductivity remains a puzzle. There is no doubt, however, that the pairing has *d*-wave symmetry. More recently, in 2006, superconductivity was discovered in LaOFeP [27], and many other iron-based superconductors were quickly identified. Once more, a different class of superconductors, iron pnictides, had been *experimentally* found, and new puzzles have to be solved; within iron pnictides a T_c as high as 57 K has been reached.

The lesson that emerges is that a superconducting state can manifest itself in very different systems, ranging from superfluid ^3He , to MgB_2 , high-temperature superconducting cuprates, and neutron stars. While the phenomenon itself is in all cases similar, its microscopic origin, i.e., the lower layer of the theory, varies strongly from case to case. The challenge is to identify in each case the proper connection between these layers of theory.

3 Overview of the school

This year's school aims to give a broad introduction to the physics of emergent phenomena in condensed matter and to the modern approaches to dealing with them. We focus primarily on the two paradigmatic manifestations of emergence that we have just discussed, magnetism and superconductivity. In order to understand these phenomena, we start with the fundamentals. The lecture of Bob Jones discusses density-functional theory from a historical perspective, stressing the aspects relevant to the study of emergence. The lecture of Erik Koch extends the scope to many-electron states, from introducing the formalism of second quantization to discussing the Mott- and BCS-states. The fundamental aspects of magnetism and exchange phenomena in a model context are presented in the lecture of Eva Pavarini. Robert Eder then introduces the variational cluster approximation to the spectral properties of the Hubbard model, the *drosophila* of strong correlation physics, using a self-energy functional.

Reflecting our focus on magnetism, a group of lectures is dedicated to magnetism in real materials and to numerical methods to solve complex spin models. The lecture of Sasha Lichtenstein retraces the path from Stoner to Hubbard models of magnetism, emphasizing modern DMFT-based approaches to understanding real materials. Treating extended magnetic systems requires highly efficient methods for successful finite-size extrapolations. Werner Krauth introduces the Monte Carlo approach and discusses methods for determining and reducing correlation times. The lecture of Stefan Wessel shows how to use Monte Carlo techniques for simulating quantum spin models. Turning to finite systems, Jürgen Schnack illustrates the state-of-the-art in describing and designing molecular magnets, intriguing systems that could become crucial building blocks for future quantum computers.

A school on emergent phenomena in condensed matter systems would not be complete without the view from experiment. Bridging magnetism and superconductivity, Bernhard Keimer took the challenging task to cover recent advances and open problems in our understanding of the high-temperature superconducting cuprates, with a special focus on the role of spin fluctuations. The next group of lectures is dedicated to the various aspects of conventional and unconventional superconductivity, the second focus of our school. The lecture of André-Marie Tremblay illustrates theoretical progress on the theory of strongly correlated superconductivity. Warren Pickett then explains the challenges in designing real superconducting materials, highlighting some of the puzzles they pose or have posed. Two lectures are dedicated to the theory of conventional superconductors. Rolf Heid discusses the mechanism of conventional superconductivity and shows how to calculate the electron-phonon coupling *ab initio* using density functional perturbation theory. These results are the input to Eliashberg theory, which is introduced in the lecture of Giovanni Umrinario. The case of superfluidity is discussed in the lecture of David Ceperley, introducing the path-integral picture of degenerate quantum systems.

The final group of lectures focuses on wave function based methods. Shiwei Zhang shows us how to study models and real materials using the auxiliary-field quantum Monte Carlo approach. Ulrich Schollwöck gives an introduction to the density-matrix renormalization group approach, while Jens Eisert explains how to analyze ground states using concepts of quantum information.

References

- [1] G.H. Lewes: *Problems of Life and Mind*,
(Kegan Paul, Trench, Turbner & Co, London, 1875)
- [2] J.S. Mill: *System of Logic* (1843)
(London: Longmans, Green, Reader, and Dyer. 8th ed., 1872)
- [3] R.B. Laughlin, and D. Pines,
Proceedings of the National Academy of Sciences (USA) **97**, 28 (2000)
- [4] P.W. Anderson, Science **177**, 393 (1972)
- [5] P.W. Anderson: *More and Different* (World Scientific, 2011)
- [6] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [7] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [8] K. Rajagopal and F. Wilczek: *The Condensed Matter Physics of QCD*,
in M. Shifman (ed.): *At the Frontiers of Particle Physics*, Vol. 3 (World Scientific, 2001)
- [9] H.J. Lipkin: *Quantum Mechanics: New Approaches to Selected Topics*
(Dover Publications, 2001)
- [10] F. Diacu, *The solution of the N-body problem*,
in *The Mathematical Intelligencer* **18**, Springer Verlag, New York (1996)
- [11] E.P. Wigner and F. Seitz, Solid State Physics **1**, 97 (1955)
- [12] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999)
- [13] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and O.K. Andersen,
Phys. Rev. Lett. **92**, 176403 (2004)
- [14] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt:
The LDA+DMFT approach to strongly correlated materials,
Reihe Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich, 2011)
<http://www.cond-mat.de/events/correl11>
- [15] P. Weiss, J. Phys. Theor. Appl. **6**, 661 (1907)
- [16] W. Heisenberg, Z. Phys. **49**, 619 (1928)
- [17] L. Néel, Ann. de Phys. **17**, 5 (1932); Ann. de Phys. **5**, 232 (1936);
C.R. Acad. Sc. **203**, 304 (1936)
- [18] C.G. Shull and J.S. Smart, Phys. Rev. **76**, 1256 (1949)

-
- [19] P.W. Higgs, Phys. Rev. Lett. **13**, 508 (1964)
- [20] More details can be found in the lecture of E. Pavarini.
- [21] H. Kamerlingh Onnes, Comm. Phys. Lab. Univ. Leiden, Vol. **119**, **120**, **122** & **124** (1911)
- [22] W. Meissner and R. Ochsenfeld, Naturwissenschaften **21**, 787 (1933)
- [23] C.A. Reynolds, B. Serin, W.H. Wright, and L.B. Nesbitt, Phys. Rev. **78**, 487 (1950);
E. Maxwell, Phys. Rev. **78**, 487 (1950)
- [24] L.N. Cooper, Phys. Rev. **104**, 1189 (1956)
- [25] J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, and J. Akimitsu,
Nature **410**, 63-64 (2001)
- [26] J. G. Bednorz and K. A. Müller, Z. Phys. B **64**, 189 (1986)
- [27] Y. Kamihara, H. Hiramatsu, M. Hirano, R. Kawamura, H. Yanagi,
T. Kamiya, and H. Hosono, J. Am. Chem. Soc. **128** 10012 (2006)

1 Density Functional Theory for Emergents

Robert O. Jones

Peter-Grünberg-Institut PGI-1 and

German Research School for Simulation Sciences

52425 Jülich, Germany

Contents

1	Introduction	2
2	The density as basic variable	3
3	An “approximate practical method”	5
4	Density functional formalism	7
4.1	Single-particle description of a many-electron system.	7
4.2	Exchange-correlation energy E_{xc} and the xc-hole	9
5	DF theory to 1990	11
5.1	Condensed matter	11
5.2	Chemistry	12
5.3	Progress to 1990	14
6	After the breakthrough	15
6.1	Progress and problems	15
6.2	An application	18
7	Summary and outlook	18

1 Introduction

The “reductionist” view of science implies focus on simpler and simpler causes and finally on the ultimate constituents of matter. This belief—that our understanding of Nature will follow if we solve this “fundamental” problem—has a vigorous opponent in Philip Anderson, who has emphasized the “emergent” view over many years [1, 2]:

“What really is the problem is called ‘broken symmetry’: the underlying laws have a certain symmetry and simplicity that is not manifest in the consequences of these laws. To put it crudely, a simple atom of gold cannot be shiny and yellow and conduct electricity: Metallicity is a property with meaning only for a macroscopic sample. ...”

Qualitatively different phenomena can occur as the level of complexity increases, and this school is devoted to understanding some of them, particularly new features that arise in “strongly correlated” systems.

A practical definition of “strongly correlated” systems covers those that are not described well by density functional (DF) theory. Many seminars and publications on “strongly correlated” mention at the outset the widespread use of density functional (DF) theory in materials science and chemistry and the physical insight that often results. The second sentence, however, often lists the systems where DF results are disastrous (an insulator is found to conduct, almost anything to do with rare earth elements, ...), emphasizing the importance of describing strongly correlated materials correctly.¹ DF theory is nevertheless an essential part of this workshop. The approach is used widely in materials science and chemistry and provides useful results for many systems for which the exact wave function cannot be computed. We should have a feel for the areas where physical insight can be obtained and *why* approximations used in DF calculations can give sensible answers far from their regions of obvious validity.

The origins of DF theory go back to the early years of quantum mechanics in the late 1920’s. Thomas [4] and Fermi [5] recognized the electron density as a basic variable, and Dirac [6] showed already in 1930 that the state of an atom can be determined completely by its density; it is not necessary to specify the wave function. We follow here the history of density-related methods to the single-particle equations of Kohn and Sham in 1965 and beyond. In its modern form, the DF formalism shows that ground state properties of a system of electrons in an external field can be determined from a knowledge of the *density distribution* $n(r)$ alone. Much of the work in materials science and chemistry focuses on a property for which DF calculations are particularly valuable: the total energy E of a system of electrons in the presence of ions located at R_I , which determines structural and cohesive properties.

Accurate calculations of the entire energy surface $E(R_I)$ are possible only for systems with very few atoms, and this function generally has vast numbers of maxima and minima at unknown locations. The lowest energy, however, corresponds to the ground state *structure*, and paths

¹An example can be found in the Preface of the 2012 Autumn School [3]: “Density functional theory (DFT) is considered the Standard Model of solid state physics. The state-of-the-art approximations to DFT, the local-density approximation (LDA) or its simple extensions, fail, however, even qualitatively, for strongly correlated systems.”

between minima are essential to our studies of chemical reactions, including their activation energies. When I read the autobiography of Francis Crick [7], I was taken by his observation

“If you want to study *function*, study *structure*.”

This may be self-evident to molecular biologists, but it is also true in many other areas. The DF approach allows us to calculate $E(R_1)$, and hence the structure and many related properties, without using experimental input. If you are more interested in “real materials” than in mathematical models, this is a crucial advantage for strongly correlated materials as well.

Olle Gunnarsson and I reviewed the density functional formalism, its history, and its prospects in 1989 [8], and I reread the original literature some years ago. My changed perspective is reflected here, where I trace DF history from the late 1920’s to the present day.

2 The density as basic variable

The recent books by Gino Segrè [9] and Graham Farmelo [10] give fascinating accounts of the development of quantum mechanics in the years following 1926. Methods for finding approximate solutions of the Schrödinger equations followed soon after the equations were published and have had a profound effect on chemistry and condensed matter physics ever since.

The “Hartree approximation” to the many-electron wave function is a product of single-particle functions [11],

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots) = \psi_1(\mathbf{r}_1) \dots \psi_N(\mathbf{r}_N) \quad (1)$$

where each $\psi_i(\mathbf{r}_i)$ satisfies a one-electron Schrödinger equation with a potential term arising from the average field of the other electrons. Hartree [11] introduced the idea of a “self-consistent field”, with specific reference to the core and valence electrons, but his papers do not mention the approximation (1). Slater [12] and Fock [13] recognized immediately that the product wave function (1) in conjunction with the variational principle led to a generalization of the method that would apply to systems more complex than atoms. They showed that replacing (1) by a determinant of such functions [12, 13] led to equations that were not much more complicated than those of Hartree, while satisfying the Pauli exclusion principle. These determinantal functions, which had been used in discussions of atoms [14] and ferromagnetism [15], are known today as “Slater determinants”, and the resulting “Hartree-Fock equations” have formed the basis of most discussions of atomic and molecular structure since.

In 1929 Dirac wrote [16]:

“The general theory of quantum mechanics is now almost complete, ... The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.”

I cannot think of a better short description of density functional theory than an “approximate practical method of applying quantum mechanics” to explain complex systems.

Dirac [16] also sought to improve the model of Thomas [4] and Fermi [5] for calculating atomic properties based purely on the electron density $n(r)$. In the first “density functional theory”, Thomas and Fermi assumed that the electrons form a homogeneous electron gas satisfying Fermi statistics and the kinetic energy has a simple dependence on the density $n(r)$. The TF equations are:

$$\frac{5}{3}C_k n(\mathbf{r})^{\frac{2}{3}} + e^2 \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{\text{ext}}(\mathbf{r}) + \lambda = 0, \quad (2)$$

where $C_k = 3\hbar^2(3\pi^2)^{\frac{2}{3}}/(10m)$, V_{ext} is the external potential, and λ is the Lagrange multiplier related to the constraint of constant particle number. Dirac noted the necessity of incorporating “exchange” phenomena, as in the Hartree-Fock approach [16], and he included these effects in the “Thomas atom” [6] by means of the potential

$$V_{\text{x}}^{\text{Dirac}} = -\left(\frac{1}{\pi}\right)(3\pi^2 n(\mathbf{r}))^{\frac{1}{3}}. \quad (3)$$

This term was derived for a homogeneous gas of density n and should be valid for weak spatial variations of $n(\mathbf{r})$.² The modified TF equation is often referred to as the “Thomas-Fermi-Dirac” equation.

The Thomas-Fermi method and its extensions give rough descriptions of the charge density and the electrostatic potential of atoms, and its mathematical properties have attracted considerable attention [17, 18]. However, it has severe deficiencies. The charge density is infinite at the nucleus and decays as r^{-6} , not exponentially, far from it. Teller [19] and others also showed that TF theory does not bind atoms to form molecules or solids, which rules out its use in chemistry or materials science. There is also no shell structure in the TF atom, so that the periodic variation of many properties with changing atomic number Z cannot be reproduced, nor can ferromagnetism [8]. Moreover, atoms *shrink* with increasing Z (as $Z^{-1/3}$) [20].

One point made by Dirac [6], however, has been emphasized by many advocates of the DF method over the years, even if we were unaware of his words of over 80 years ago:

“ Each three-dimensional wave function will give rise to a certain electric density. This electric density is really a matrix, like all dynamical variables in the quantum theory. By adding the electric densities from all the wave functions we can obtain the total electric density for the atom. If we adopt the equations of the self-consistent field as amended for exchange, then this total electric density (the matrix) has one important property, namely, if the value of the total electric density at any time is given, then its value at any later time is determined by the equations of motion. This means that *the whole state of the atom is completely determined by this electric density; it is not necessary to specify the individual three-dimensional wave functions that make up the total electric density*. Thus one can deal with any number of electrons by working with just one matrix density function.”

²The exchange energy in a homogeneous electron gas had been derived by Bloch [15] in 1929

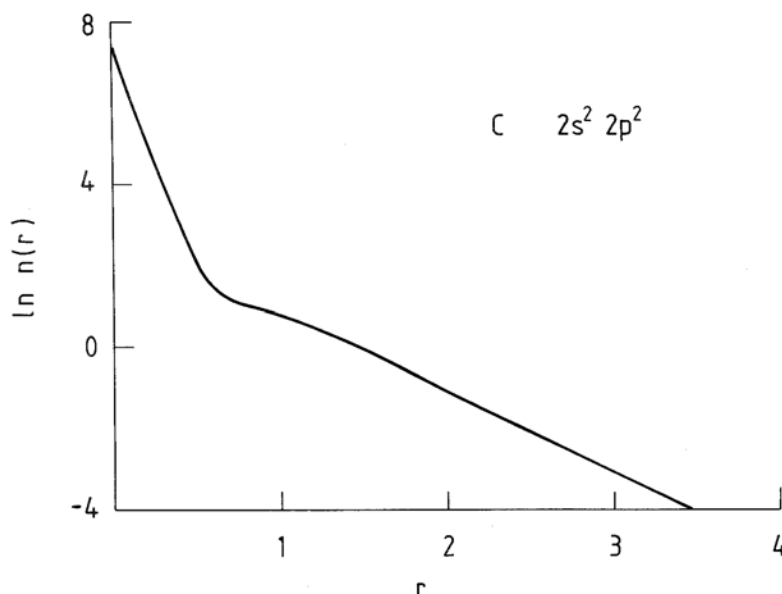


Fig. 1: *Logarithm of spherical average of density in ground state of C atom as a function of the distance from the nucleus (atomic units) [8].*

The italics are in the original. The derivation is based on the “self-consistent field” or Hartree-Fock approximation, but the observation that the density follows the equations of motion is much in the spirit of Ehrenfest’s theorem [21], which has wider validity. Ehrenfest had proved in 1927 what I have seen referred to as the “time-dependent Hellmann-Feynman theorem”, namely that the acceleration of a quantum wave packet that does not spread satisfied Newton’s equations of motion.

The central role played by the density means that we must understand its nature in real systems. Figure 1 shows that the spherically averaged density in the ground state of the carbon atom falls monotonically from the nucleus and does not show the radial oscillations that occur if we plot $r^2n(r)$. The charge density in small molecules is also relatively featureless, with maxima at the nuclei, saddle points along the bonds, and a generally monotonic decay from both. The electron density in molecules and solids also shows relatively small departures from the overlapped densities of the constituent atoms. Energy differences, including binding, ionization, and cohesive energies, are the focus of much DF work and result from subtle changes in relatively featureless density distributions. It really is amazing that this suffices to determine ground state properties.

3 An “approximate practical method”

The basis of a quantum theory of atoms, molecules, and solids was in place at the beginning of the 1930’s. Linear combinations of atomic orbitals formed molecular orbitals, from which determinantal functions could be constructed, and linear combinations of determinants (“configuration interaction”) would provide approximations to the complete wave function. Dirac had noted already, however, that this procedure could not be implemented in practice, so that

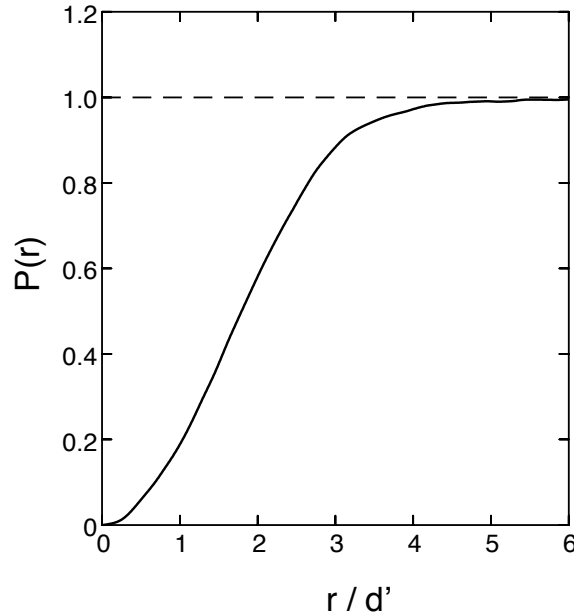


Fig. 2: Probability that electrons in Na metal with parallel spins are r/d' apart ($d'^3 = V_0/(3\pi^2)$, V_0 is the atomic volume). After Wigner and Seitz [22].

approximations are essential. Furthermore, numerical techniques for solving the Schrödinger equation in extended systems needed to be developed.

Wigner and Seitz [22] developed a method for treating the self-consistent problems in crystals, and the “Wigner-Seitz cell” is known to all condensed matter physicists. The first application to metallic sodium used a pseudopotential for the Na ion, and calculations of the lattice constant, cohesive energy, and compressibility gave satisfactory results. Of particular interest for our purposes, however, is the calculation of the probability of finding electrons with parallel spins a distance r apart (Fig. 2). This function obtains its half-value for $r = 1.79 d'$ or $0.460 d$ for a body-centred cubic lattice with cube edge d , which is close to the radius of the “Wigner-Seitz sphere” $(\frac{3}{8\pi})^{\frac{1}{3}} d = 0.492 d$. The exclusion principle means then that two electrons with parallel spins will very rarely be at the same ion. This argument does not depend significantly on the potential and should apply to a Fermi gas subject to periodic boundary conditions [22]. The corresponding curves for spin up and spin down electrons, as well as for both spins combined, were discussed in the 1934 review article of Slater [23].

The picture that results is simple and appealing: the exclusion principle means that an electron with a given spin produces a surrounding region where there is a deficiency of charge of the same spin. This region contains unit charge and is referred to as the “Fermi” [22] or “exchange” hole [24]. In the Hartree-Fock scheme, the exchange hole is different for each electronic function, but Slater [24] developed a simplified “exchange potential” that depended only on the density:

$$V_x^{\text{Slater}} = -\left(\frac{3}{2\pi}\right)(3\pi^2 n(r))^{\frac{1}{3}}. \quad (4)$$

The Slater approximation (4) was proposed at the time that electronic computers were becoming

available for electronic structure calculations and proved to be very useful in practice. Methods for solving the Schrödinger equation had been developed around this time, including the augmented plane wave (APW) [25] and Korringa-Kohn-Rostoker approaches [26, 27].

The exchange potential of Slater (4) is 3/2 times that derived by Dirac and Bloch (3) for a homogeneous electron gas, but Slater [28] pointed out that an effective potential proportional to the cube root of the density could be obtained by arguments based on the exchange hole that do not depend on the free electron gas arguments used in the original derivation [24]. The exchange hole discussed above for a spin up electron contains a single electron. If we assume that it can be approximated by a sphere of radius R_\uparrow , then

$$\left(\frac{4\pi}{3}\right)R_\uparrow^3 n_\uparrow = 1 \quad ; \quad R_\uparrow = \left(\frac{3}{4\pi n_\uparrow}\right)^{\frac{1}{3}} \quad (5)$$

where n_\uparrow is the density of spin up electrons. Since the electrostatic potential at the centre of such a spherical charge is proportional to $1/R_\uparrow$, the exchange potential will be proportional to $n_\uparrow^{\frac{1}{3}}$. This argument was used by Slater to counter a (still widespread) misconception that local density approximations are only appropriate if the electron density is nearly homogeneous.

In 1954, Gáspár [29] questioned the prefactor of the effective exchange potential (Eq. 4). If one varies the spin orbitals to minimize the total energy in the Thomas-Fermi-Dirac form, one obtains a coefficient just $\frac{2}{3}$ as large. Gáspár applied this approximation to the Cu^+ ion and found good agreement with Hartree-Fock eigenfunctions and eigenvalues. Slater noted that Gáspár's method was "more reasonable than mine" [30], but the larger value was used in most calculations in the following years.

4 Density functional formalism

The variational principle on the energy was the basis of the formulation of the density functional formalism given by Hohenberg and Kohn [31]. The ground state (GS) properties of a system of electrons in an external field can be expressed as functionals of the GS electron density, i.e. they are determined by a knowledge of the density alone. The total energy E can be expressed in terms of such a functional, and $E[n]$ satisfies a variational principle. These theorems were proved by Hohenberg and Kohn [31] for densities that can be derived from the ground state of some external potential V_{eff} ("V-representable"). A simpler and more general proof for ("N-representable") densities that can be derived from some antisymmetric wave function was given by Levy [32, 33]. Of course, these proofs do not provide practical prescriptions for writing the functional relationship between energy E and density n .

4.1 Single-particle description of a many-electron system.

The task of finding good approximations to the energy functional $E(n)$ is simplified greatly if we use the decomposition introduced by Kohn and Sham [34],

$$E[n] = T_0[n] + \int d\mathbf{r} n(\mathbf{r}) \left(V_{\text{ext}}(\mathbf{r}) + \frac{1}{2}\Phi(\mathbf{r}) \right) + E_{\text{xc}}[n] . \quad (6)$$

T_0 is the kinetic energy that a system with density n would have if there were no electron-electron interactions, Φ is the classical Coulomb potential for electrons, and E_{xc} defines the exchange-correlation energy. T_0 is not the true kinetic energy T , but it is of comparable magnitude and is treated here without approximation. This removes many of the deficiencies of the Thomas-Fermi approach, such as the lack of a shell structure of atoms or the absence of chemical bonding in molecules and solids. In the expression (6) all terms other than the exchange-correlation energy E_{xc} can be evaluated exactly, so that approximations for this term are crucial in density functional applications.

The variational principle applied to (6) yields

$$\frac{\delta E[n]}{\delta n(\mathbf{r})} = \frac{\delta T_0}{\delta n(\mathbf{r})} + V_{\text{ext}}(\mathbf{r}) + \Phi(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} = \mu , \quad (7)$$

where μ is the Lagrange multiplier associated with the requirement of constant particle number. If we compare this with the corresponding equation for a system with an effective potential $V(\mathbf{r})$ but *without* electron-electron interactions,

$$\frac{\delta E[n]}{\delta n(\mathbf{r})} = \frac{\delta T_0}{\delta n(\mathbf{r})} + V(\mathbf{r}) = \mu , \quad (8)$$

we see that the mathematical problems are identical, provided that

$$V(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \Phi(\mathbf{r}) + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} . \quad (9)$$

The solution of (Eq. 8) can be found by solving the Schrödinger equation for *non-interacting* particles,

$$\left(-\frac{1}{2}\nabla^2 + V(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) , \quad (10)$$

yielding

$$n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2 \quad (11)$$

The condition (9) can be satisfied in a self-consistent procedure.

The solution of this system of equations leads to the energy and density of the lowest state, and all quantities derivable from them. The formalism can be generalized to the lowest state with a given symmetry [35]. Instead of seeking these quantities by determining the wave function of the system of interacting electrons, the DF method reduces the problem to the solution of a single-particle equation of Hartree form. In contrast to the Hartree-Fock potential,

$$V_{\text{HF}} \psi(\mathbf{r}) = \int d\mathbf{r}' V_{\text{HF}}(\mathbf{r}, \mathbf{r}') \psi(\mathbf{r}') , \quad (12)$$

the effective potential, $V(\mathbf{r})$ is *local*.

The numerical advantages of solving the Kohn-Sham equations [34] are obvious. Efficient methods exist for solving single-particle Schrödinger-like equations with a local effective potential, and there is no restriction to small systems. With a local approximation to E_{xc} , the

equations can be solved as readily as the Hartree equations. Unlike the Thomas-Fermi method, where the large kinetic energy term is approximated, the valence kinetic energy and the core-valence and valence-valence electrostatic interactions are treated exactly. However, E_{xc} is the difference between the *exact* energy and terms we can evaluate exactly, and approximations are unavoidable.

4.2 Exchange-correlation energy E_{xc} and the xc-hole

Kohn and Sham [34] proposed using the “local density (LD) approximation”

$$E_{xc}^{LD} = \int d\mathbf{r} n(\mathbf{r}) \varepsilon_{xc}[n(\mathbf{r})] , \quad (13)$$

where $\varepsilon_{xc}[n]$ is the exchange and correlation energy per particle of a homogeneous electron gas with density n . This approximation is exact in the limits of slowly varying densities and very high densities. The authors noted that this approximation “has no validity” at the “surface” of atoms and in the overlap regions of molecules and concluded [34]:

“We do not expect an accurate description of chemical bonding.”

The generalization to spin-polarized systems is

$$E_{xc}^{LSD} = \int d\mathbf{r} n(\mathbf{r}) \varepsilon_{xc}[n_{\uparrow}(\mathbf{r}), n_{\downarrow}(\mathbf{r})] , \quad (14)$$

where $\varepsilon_{xc}[n_{\uparrow}, n_{\downarrow}]$ is the exchange and correlation energy per particle of a homogeneous, spin-polarized electron gas with spin-up and spin-down densities n_{\uparrow} and n_{\downarrow} , respectively.³ The “ $X\alpha$ ” approximation

$$E_x^{X\alpha} = -\frac{3}{2}\alpha C \int d\mathbf{r} \left((n_{\uparrow}(\mathbf{r}))^{4/3} + (n_{\downarrow}(\mathbf{r}))^{4/3} \right) , \quad (15)$$

where $C = 3(3/4\pi)^{1/3}$ was used in numerous calculations in the late 1960’s and 1970’s. The α -dependence of energy differences for a given atom or molecule is weak for values near 2/3, the value of Dirac [6], Bloch [15], Gáspár [29] and Kohn and Sham [34]. We have noted that the electron density in molecules and solids is generally far from that of a homogeneous electron gas, and the validity of calculations based on properties of a gas of constant density has often been questioned. We now discuss some general properties of E_{xc} using arguments closely related to the “exchange hole” picture of Wigner and Seitz [22] and Slater [24, 28].

The crucial simplification in the density functional scheme is the relationship between the interacting system, whose energy and density we seek, and the fictitious, non-interacting system for which we solve (Eq. 10, 11). This can be studied by considering the interaction $\lambda/|\mathbf{r} - \mathbf{r}'|$ and varying λ from 0 (non-interacting system) to 1 (physical system). This is done in the presence of an external potential V_{λ} , [36] such that the ground state of the Hamiltonian

$$H_{\lambda} = -\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_{\lambda} + \lambda V_{ee} \quad (16)$$

³The calculation by Bloch [15] in 1929 of ferromagnetism in a free-electron model of a metal was the first where the exchange energy was expressed as the sum of terms proportional to $n_{\uparrow}^{4/3}$ and $n_{\downarrow}^{4/3}$.

has density $n(\mathbf{r})$ for all λ . The exchange-correlation energy of the interacting system can then be expressed as an integral over the coupling constant λ [35]:

$$E_{\text{xc}} = \frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) \int d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|} n_{\text{xc}}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) , \quad (17)$$

with

$$n_{\text{xc}}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) \equiv n(\mathbf{r}') \int_0^1 d\lambda \left(g(\mathbf{r}, \mathbf{r}', \lambda) - 1 \right) . \quad (18)$$

The function $g(\mathbf{r}, \mathbf{r}', \lambda)$ is the pair correlation function of the system with density $n(\mathbf{r})$ and Coulomb interaction λV_{ee} . The exchange-correlation hole, n_{xc} , describes the fact that an electron at point \mathbf{r} reduces the probability of finding one at \mathbf{r}' , and E_{xc} is simply the energy resulting from the interaction between an electron and its exchange-correlation hole. This is a straightforward generalization of the work of Wigner and Seitz [22] and Slater [24] discussed above. Second, the isotropic nature of the Coulomb interaction V_{ee} has important consequences. A variable substitution $\mathbf{R} \equiv \mathbf{r}' - \mathbf{r}$ in (17) yields

$$E_{\text{xc}} = \frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) \int_0^\infty dR R^2 \frac{1}{R} \int d\Omega n_{\text{xc}}(\mathbf{r}, \mathbf{R}) . \quad (19)$$

Equation (19) shows that the xc-energy depends only on the spherical average of $n_{\text{xc}}(\mathbf{r}, \mathbf{R})$, so that approximations for E_{xc} can give an *exact* value, even if the description of the non-spherical parts of n_{xc} is arbitrarily inaccurate. Thirdly, the definition of the pair-correlation function leads to a sum-rule requiring that the xc-hole contains one electron, i.e. for all \mathbf{r} ,

$$\int d\mathbf{r}' n_{\text{xc}}(\mathbf{r}, \mathbf{r}' - \mathbf{r}) = -1 . \quad (20)$$

This means that we can consider $-n_{\text{xc}}(\mathbf{r}, \mathbf{r}' - \mathbf{r})$ as a normalized weight factor, and define locally the radius of the xc-hole,

$$\left\langle \frac{1}{\mathbf{R}} \right\rangle_{\mathbf{r}} = - \int d\mathbf{r} \frac{n_{\text{xc}}(\mathbf{r}, \mathbf{R})}{|\mathbf{R}|} . \quad (21)$$

This leads to

$$E_{\text{xc}} = -\frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) \left\langle \frac{1}{\mathbf{R}} \right\rangle_{\mathbf{r}} . \quad (22)$$

Provided Equation (20) is satisfied, E_{xc} is determined by the first moment of a function whose second moment we know exactly and depends only weakly on the details of n_{xc} [35]. Approximations to E_{xc} can then lead to good total energies (and structures), even if the details of the exchange-correlation hole are described very poorly. This is shown in Figure 3, where the exchange hole in a nitrogen atom is shown for a representative value of \mathbf{r} for both the local density and exact (Hartree-Fock) cases. The holes are qualitatively different: The LD hole is spherically symmetric and centred on the electron, while the exact hole has a large weight at the nucleus and is very asymmetric. Nevertheless, the spherical averages are very similar, and the exchange energies differ by only a few percent.

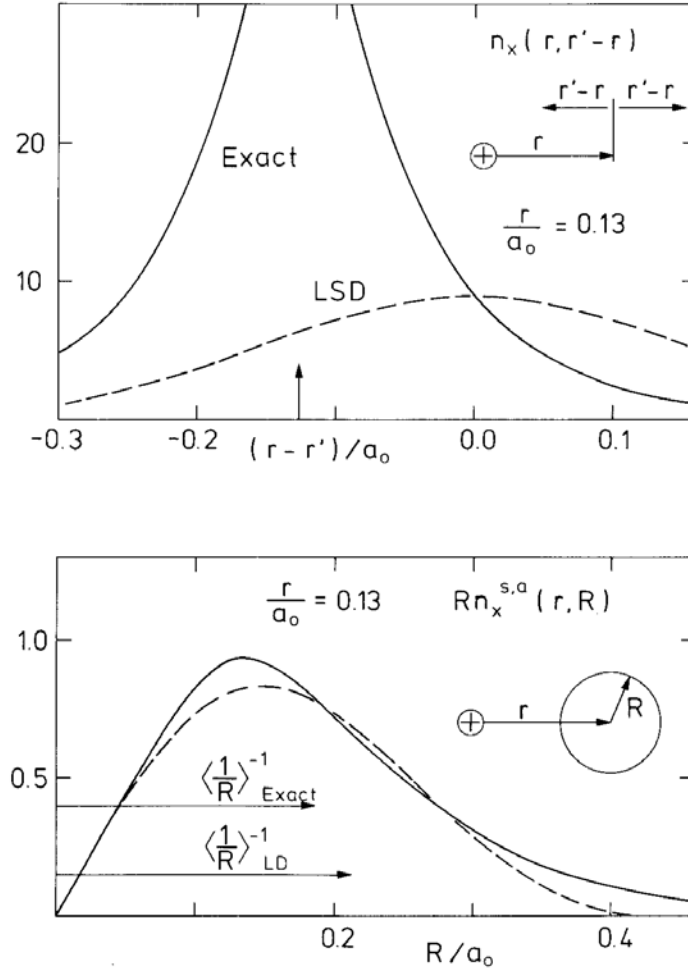


Fig. 3: Exact (solid) and LSD (dashed) exchange holes $n_{xc}(\mathbf{r}, \mathbf{r}' - \mathbf{r})$ for spin up electron in N atom for $r = 0.13$ a.u. Upper: hole along line through nucleus (arrow) and electron ($\mathbf{r} - \mathbf{r}' = 0$). Lower: spherical averages of holes, and $\langle 1/R \rangle$ (21) [8].

5 DF theory to 1990

5.1 Condensed matter

Condensed matter physicists were generally pleased to have justification for the “local density” calculations they had been performing for years, and numerous electronic structure theorists moved seamlessly from performing “ $X\alpha$ ” or “Hartree-Fock-Slater” calculations into the density functional world. However, Fig. 4 shows that there was remarkably little impact of DF calculations prior to 1990. Volker Heine, a prominent condensed matter theorist, looked back on the 1960’s in this way [37]:

“ Of course at the beginning of the 1960s the big event was the Kohn Hohenberg Sham reformulation of quantum mechanics in terms of density functional theory (DFT). Well, we recognize it now as a big event, but it did not seem so at the time. That was the second big mistake of my life, not to see its importance, but then neither did the authors judging from the talks they gave, nor anyone else. Did you ever wonder why they never did any calculations with it?”

There were also prominent critics of density functional and related computational techniques, and one of the best known solid state theoreticians, Philip Anderson, made devastating comments in 1980 [38]:

“ There is a school which essentially accepts the idea that nothing further is to be learned in terms of genuine fundamentals and all that is left for us to do is calculate. . . . One is left, in order to explain any phenomenon occurring in ordinary matter, only with the problem of doing sufficiently accurate calculations. This is then the idea that I call “The Great Solid State Physics Dream Machine” . . . This attitude is closely associated with work in a second field called quantum chemistry.”

Anderson associated the “Dream Machine” with the name of John Slater and described the DF method as a “simplified rather mechanical kind of apparatus” that “shows disturbing signs of become a victim of the “Dream Machine” syndrome” [38]. While noting that DF calculations can be particularly valuable in some contexts, he continued:

“ . . . a great deal of the physics is concealed inside the machinery of the technique, and that very often once one has the answers that these techniques provide, one is not exactly clear what the source of these answers is. In other words the better the machinery, the more likely it is to conceal the workings of nature, in the sense that it simply gives you the experimental answer without telling you why the experimental answer is true.”

While some may find these words a little harsh, his comments did apply to some electronic structure calculations at the time. They may indeed have had prophetic character, as I discuss in Sec. 7. The increasing availability of computing resources made possible calculations that had previously been inaccessible, and not all users of the method were critical of the approximations involved.

5.2 Chemistry

It took many years for DF calculations to be taken seriously by most chemists, and the reasons were often convincing: (1) Unlike the TF theory, the Kohn-Sham expression for the energy is not really a “functional” of the density, since the kinetic energy term is treated exactly and is defined by an effective potential that leads to the density, (2) the original functional of Hohenberg and Kohn is not even *defined* for all n , because not all densities can be derived from the ground state of some single-particle potential [33], (3) approximations to the exchange-correlation energy are unavoidable, and their usefulness can be assessed only by trying them out, and (4) there is no *systematic* way to approach the exact solution of the Schrödinger equation and, of course, the exact energy.

This last point was emphasized by many. In principle, the Hartree-Fock method could be extended to multiple determinants (“configuration interaction”) and, coupled with a large basis set, lead to the exact wave function and all properties obtainable from it. This is a very attractive proposition, and the dramatic improvements in computing power (three orders of magnitude per

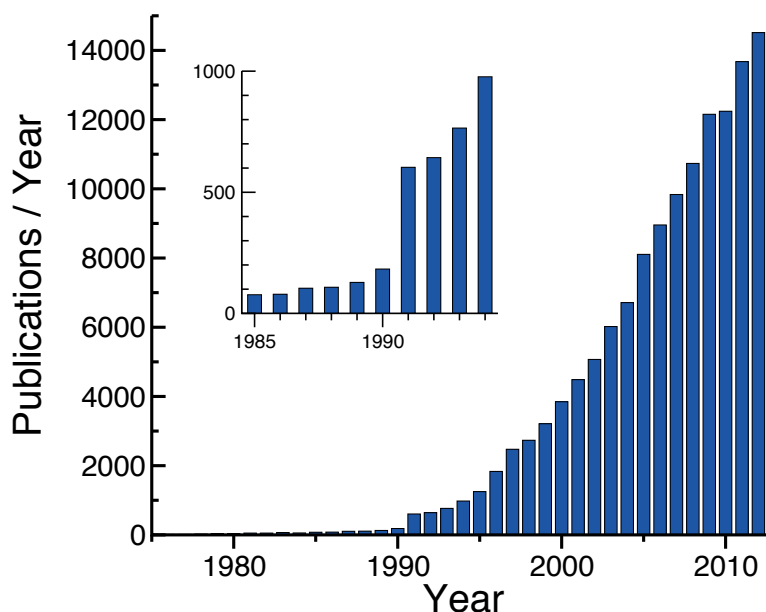


Fig. 4: Number of publications per year (1975-2012) on topics “density functional” or “DFT”, according to Web of Knowledge (May 2013). Inset shows data near 1990 on an expanded scale [39].

decade) might make the reservations of Dirac [16] less formidable. It was often emphasized that solutions of the Schrödinger equation led to the “right answer for the right reason.” Nevertheless, obtaining numerically exact total energies from calculations of the wave function remains a major challenge to this day, and it is not surprising that several groups looked at alternatives. Hartree-Fock-Slater calculations on small molecules were carried out in the early 1970’s, particularly by Evert Jan Baerends and collaborators in Amsterdam, and some of the first DF calculations on small molecules were performed by Olle Gunnarsson [35]. John Harris and I had not expected that the local density approximations would give reasonable results for molecules, and we (with Olle) developed a full-potential LMTO code for small molecules and clusters [40]. These calculations led to good geometries and reasonable binding energies in most cases. In spite of the shortcomings of the local density description of E_{xc} , it was now possible to perform calculations without adjustable parameters on families of molecules and small clusters that had previously been inaccessible. I was almost unprepared for so many really exciting results, my own examples including the trends in the binding energies of group 2 dimers [41, 42] and the structures of small phosphorus clusters [43]. Most condensed matter physicists were either not surprised or not interested, but theoretical chemists remained sceptical or critical, and this situation continued throughout the 1980’s and into the 1990’s.

The Seventh International Congress of Quantum Chemistry, held in Menton, France, from 2-5 July 1991, marks for me a major turning point in the fortunes of DF methods in chemistry. Density-related methods were discussed in detail, and communication between their proponents and the sceptics improved. Becke described his development of a non-local exchange functional that promised improvements over local approximations [44], and this approximation was tested

for the atomization energies of small molecules immediately after the meeting. Many will have been surprised by the results [45]:

“ In summary, these initial results indicate that DFT is a promising means of obtaining quantum mechanical atomization energies; here, the DFT methods B-VWN and B-LYP outperformed correlated ab initio methods, which are computationally more expensive.”

and [46]

“ The density functional vibration frequencies compare favorably with the ab initio results, while for atomization energies two of the DFT methods give excellent agreement with experiment and are clearly superior to all other methods considered.”

The ab initio methods mentioned were Hartree-Fock, second order Møller-Plesset (MP2), and quadratic configuration interaction with single and double substitutions (QCISD). In addition to the growing body of results on molecules and clusters that were beyond the scope of calculations of correlated wave functions, this change in attitude by one of the most prominent theoretical chemists led to a dramatically new attitude towards the DF method in chemistry.

5.3 Progress to 1990

The number of citations to density functional theory and related topics was very small prior to 1990 and exploded thereafter (see Figure 4). However, work was already in place by 1990 that has proved to be crucial to the ultimate acceptance of the method, and I now outline some of it. More details can be found elsewhere [8, 47].

The generalizations to finite temperatures and to spin systems were carried out soon after the original work of Hohenberg and Kohn [31]. The former was provided by Mermin [48], who showed that, in a grand canonical ensemble at given temperature T and chemical potential μ , the equilibrium density is determined by the external potential V_{ext} , and the equilibrium density minimizes the grand potential. Single-particle equations can be derived for a fictitious system with kinetic energy T_0 and entropy S_0 , with E_{xc} replaced by the exchange-correlation contribution to the free energy.

The extension to spin systems [49] or an external magnetic field requires the introduction of the spin indices α of the one-electron operators $\psi_\alpha(\mathbf{r})$ and replacing V_{ext} by $V_{\text{ext}}^{\alpha\beta}(\mathbf{r})$, and the charge density $n(\mathbf{r})$ by the density matrix $\rho_{\alpha\beta}(\mathbf{r}) = \langle \Psi | \psi_\beta^\dagger(\mathbf{r}) \psi_\alpha(\mathbf{r}) | \Psi \rangle$. All ground state properties are functionals of $\rho_{\alpha\beta}$, and E is stationary with respect to variations in $\rho_{\alpha\beta}$. The expression for the energy E_{xc} is analogous to Equations (17,18). A current- and spin density functional theory of electronic systems in strong magnetic fields was formulated by Vignale and Rasolt [50]. Time-dependent density functional theory, which has proved to be valuable in discussing excited states, was described by Runge and Gross [51].

Most of the early DF calculations on small clusters and molecules used the LD and/or LSD approximations. Although the results were generally encouraging, it was soon clear that local density calculations can lead to unacceptable errors. Examples were the exchange energy difference between states with different nodal structures [52], including the s - p promotion energies

in first-row atoms, particularly O and F. Dispersion forces – the weak, non-local interactions between closed shells systems – are a particular problem for such approximations. The long-range interaction between separated atoms or molecules is absent, and yet the LD approximation *overestimates* the binding energy in many such systems, e.g. He_2 [41]. It is not surprising that new approximations were developed, and corrections involving density gradients were soon available for the correlation [53, 54] and exchange energies [44]. The semi-empirical exchange energy approximation of Becke [44] had the correct asymptotic behaviour for atoms.

The combination of DF calculations with molecular dynamics (Car-Parrinello method) [55] made simulations of bulk systems at elevated temperatures possible, and simulated annealing techniques could be used to study the energy surfaces of molecules and clusters. My 1991 article [56] showed that unexpected structures could result. An essential part of DF work prior to 1990 was, of course, the gradual generation of a data base of results for molecules and clusters.

6 After the breakthrough

There have been over 134,000 publications on the topics “density functional” and “DFT” between 1990 and May 2013 (Figure 4), and I leave detailed surveys of this vast literature to others. I mention here some aspects that should be of general interest and give an example of the possibilities provided by the combination of DF calculations with molecular dynamics.

6.1 Progress and problems

One of the first signs of growing acceptance of DF methods in chemistry was the incorporation of such calculations into popular *ab initio* program packages, with Gaussian leading the way. It seems that Michael Frisch, first author of that package, was a willing convert. At the end of a talk at the ACS National Meeting in San Francisco (13 April 1997) on “*Ab initio* calculations of vibrational circular dichroism and infrared spectra using SCF, MP2, and density functional theories for a series of molecules,” an unknown (to me) member of the audience asked:

“What about Hartree-Fock?”

I wrote his answer down without delay:

“It does not matter what you want to calculate, and it does not matter what functional you use; density functional results are always better than Hartree-Fock.”

The availability of such codes and the possibility of comparing the results of different types of calculation were important to establishing the credentials of DF calculations in chemistry.

There has been progress in all the above areas. Time-dependent DF theory has become a standard way to calculate excited states and is an option in most DF program packages. The number of publications in a year that use the Car-Parrinello method has grown nearly linearly from

almost zero in 1990 to 1400 in 2012 [39]. The combination of DF calculations for a chemically active region with classical molecular dynamics for the surrounds (the “QM/MM approach”) [57] has found applications in many systems in biology, as well as organic and solid state chemistry [58]. Classical force fields that lead to simulations with near-DF accuracy can be developed by a neural network representation of the results of (many) DF calculations on small systems [59]. There are lengthy reviews of orbital-dependent (and other) density functionals [60] and constrained density functional theory [61]. The random phase approximation (RPA) is being tested in various contexts [62, 63].

These and other developments are very welcome, but they do not hide the fact that the most contentious issue has been the development of approximations to the exchange-correlation energy that overcome the weaknesses of the local density approximations. The LD (Eq. (13)) and LSD (Eq. (14)) approximations lead to overbinding of many molecules, poor exchange energy differences if the nodal structures of the orbitals change, and the Kohn-Sham eigenvalues often underestimate measured optical band gaps. Nevertheless, calculations that used them provided insight into many physical problems, and the reasons for the errors (and ways to assess their magnitude) became clear. However, if insight is not enough and reliable *numbers* are needed, improved approximations are necessary.

The first generalized gradient approximations [44, 53, 54] did lead to better results, and hybrid functionals including exact exchange were introduced by Becke in 1993 [64]. This form of E_x has three parameters, and its combination with E_c of Lee, Yang, and Parr [54] (B3LYP) is still the most common approximation used in chemical applications [65]. Many other empirical and hybrid functionals have been developed since, with parameters usually fit to thermochemical data for particular groups of molecules. The use of experimental data for fitting functional forms is understandable [66]. The additional parameters led to improvement over the LD and LSD results, and the use of “training sets” of atomic and molecular systems to optimize the parameters improved the calculated results for particular sets of molecules [67].

An alternative path has been followed by others, particular Perdew and collaborators, who developed a sequence (“Jacob’s ladder”) of approximations without experimental input, where each “rung” built on the experience of lower level and satisfies particular physical constraints. The gradient corrected form of Perdew, Burke, and Ernzerhof [68] (PBE) incorporates the LSD form below it, and the “meta-GGA” form of Tao, Perdew, Staroverov, and Scuseria (TPSS) [69], where n_\uparrow and n_\downarrow are joined by their gradients and the kinetic energy density of the occupied Kohn-Sham orbitals, built on both. The agreement with experiment improves (and the complexity of the calculations increases) as one climbs the “ladder” [70].

Two areas have remained particular challenges for DF calculations. The first are the weak dispersion or van der Waals forces mentioned above, where there has been substantial progress during recent years. The development of a functional that changes seamlessly on going from weakly interacting units to a combined system has been a goal of many, and one successful project has been that of Langreth and coworkers [71]. Their functional incorporates results for electron gas slabs and the electron gas itself, is free of experimental input, and has been implemented in several program packages. An empirical correction to DF results has been

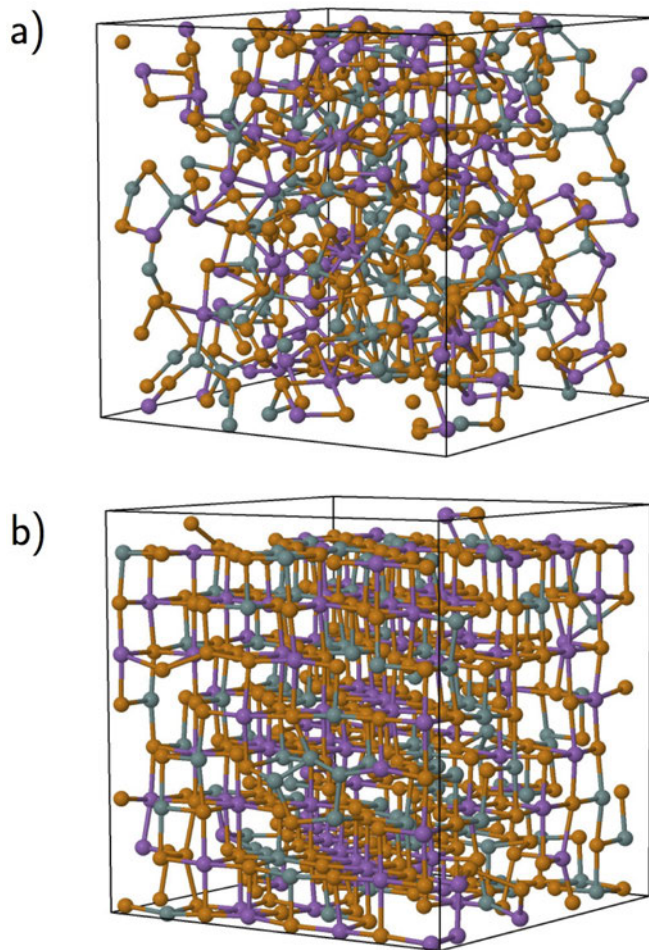


Fig. 5: Crystallization in GST alloy at 600 K. (a) Amorphous structure after 215 ps, (b) crystalline structure after 1045 ps. Green: Ge, purple: Sb, Orange: Te.

made by Grimme [72], and an alternative has been suggested by Tkatchenko and Scheffler [73]. “Strongly correlated” systems are those where the interaction energy dominates over the kinetic energy and often involve transition element or rare earth atoms. Local density approximations can give qualitatively incorrect descriptions of these materials, and the use of model Hamiltonians has been a popular way to avoid them. A common approach has been to add an on-site Coulomb repulsion (“Hubbard U ”) in the “LSD+ U ” scheme [74, 75]. The parameter U can be estimated within a DF framework [75, 76] or fit to experiment.

There are developments in the quantum Monte Carlo (QMC) studies of *interacting* electron systems that could be relevant for future DF work. The full configuration interaction (FCI) implementation of QMC has been applied recently to the homogeneous electron gas [77] and to simple solids [78]. Condensed matter scientists have much experience with periodic boundary conditions and plane wave orbital expansions, and this should aid the implementation of the method in extended systems. Another example is the reformulation of the constrained search approach in DF theory [32, 33] in terms of the density *and* the $(N - 1)$ -conditional probability density, which can be treated by ground state path integral QMC [79]. It remains to be seen

whether the computational demands usually associated with QMC can be reduced.

The terms “ab initio” and “first principles” are used differently in the “chemical” and “materials” worlds. For most chemists, the expressions means solutions of the Schrödinger equation for the system of interacting electrons (e.g. by QMC), for materials scientists it can be a DF calculation without (or even with) adjustable parameters. I carry out “density functional” calculations and describe them as such, and I am happy to use the term “ab initio” for solutions of the Schrödinger equation, as done by chemists.

6.2 An application

The results of one DF simulation in materials science using the PBE functional [68] show the scale of DF simulations that are possible today. Phase change (PC) materials are alloys of chalcogens (group 16 elements) that are ubiquitous in the world of rewritable optical storage media, examples being the digital versatile disk (DVD-RW) and Blu-ray Disc. Nanosized bits in a thin polycrystalline layer are switched reversibly and extremely rapidly between amorphous and crystalline states, and the state can be identified by changes in resistivity or optical properties. Crystallization of the amorphous bit is the rate-limiting step in the write/erase cycle, and much attention has been focused on this process. Alloys of Ge, Sb, and Te are often used in PC materials, and 460-atom simulations have been carried out at 600 K on amorphous $\text{Ge}_2\text{Sb}_2\text{Te}_5$ [80] (Fig. 5). Crystallization takes place in just over 1 ns, and it is possible to monitor changes in the distribution of the cavities, the diffusion of atoms of the different elements, and percolation of crystalline units in the sample. These calculations involve over 400,000 (!) self-consistent DF calculations of structure, energies, and forces for a 460-atom sample. The steady (and essential) improvement in numerical algorithms has played an important role, but such calculations also require computers of the highest performance class.

7 Summary and outlook

The astonishing growth of density functional calculations since 1990 resulted in the award of the 1998 Nobel Prize for Chemistry to Walter Kohn. Although he noted that “very deep problems” remain, Philip Anderson felt that this award may indicate that [81]

“the labours and controversies ... in understanding the chemical binding in materials had finally come to a resolution in favour of ‘LDA’ and the modern computer”,

The LD and LSD approximations have well documented drawbacks, and the resulting *numbers* (binding energies, band gaps, ...) should be treated with caution. However, the approximations satisfy important physical criteria, such as the sum rule on the exchange-correlation hole, and our long experience with them helps us to judge when the results may be wrong and by how much. The bonding *patterns* are correct in most cases, which is no doubt one reason why LD approximations and their modifications are still in widespread use. They make possible the simultaneous study of numerous related systems, such as families of molecules or materials, with

the computational resources needed to determine the wave function of a single much smaller system.

This pragmatic approach seems to be giving way to the search for schemes that produce better *numbers* automatically, preferably without having to worry about the nature of the system or the bonding mechanism involved. The long and growing list of approximate functionals and countless publications comparing their predictions with each other and with experiment have led to a chaotic situation. A newcomer to the field must despair of understanding why one approximation should be favoured over another or the real physical reasons behind a particular result. Are DF calculations in chemistry now following the “Dream Machine” scenario foreseen for the solid state world by Anderson in 1980? [38]. Furthermore, a comparison of the band gaps in LiH and four alkali halides, four oxides, and solid Ne and Ar (gaps between 0.2 and 20 eV) with the predictions of many popular functionals [82] showed that the identification of the “best” functional depends on the choice of statistical measure (mean error, mean absolute error, variance, ...)!

Density functional theory deserves better than to be a background justification for empirical curve fitting, which clearly implies a lack of confidence in the theory, or the development of a never ending chain of approximations seeking the “right” numbers, with less concern for their physical origin. It is a wonderful development with a long and fascinating history involving some of the best known names in physics. It may not provide precise answers to some questions using simple descriptions of the exchange-correlation energy, but its ability to outperform methods that seek exact solutions of the Schrödinger equation is not threatened. We shall continue to obtain insight into all sorts of problems that we cannot imagine today.

I end with a note of caution for the “strongly correlated” community. Few theoretical chemists thought that DF calculations were relevant to understanding the electronic structure of molecules, but local density approximations (and their modifications) have given far better results than *any-one* expected. It was shown *afterwards* (see, for example, Sec. 4.2) why approximations to E_{xc} could give good results for density distributions far from those where they are obviously valid. Perhaps DF theory has some real surprises in store for the “strongly correlated” world.

References

- [1] P.W. Anderson, *Science* **177**, 393 (1972)
- [2] P.W. Anderson, *More and Different. Notes from a thoughtful curmudgeon* (World Scientific, Singapore, 2011), pp. 134-139
- [3] E. Pavarini, E. Koch, F. Anders, and M. Jarrell (eds.):
Correlated Electrons: From Models to Materials
Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
<http://www.cond-mat.de/events/correl12>
- [4] L.H. Thomas, *Proc. Camb. Phil. Soc.* **23**, 542 (1927)
- [5] E. Fermi, *Z. Physik* **48**, 73 (1928)
- [6] P.A.M. Dirac, *Proc. Camb. Phil. Soc.* **26**, 376 (1930)
- [7] F.H.C. Crick, *What mad pursuit* (Penguin, London, 1988), p. 150
- [8] R.O. Jones, O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989)
- [9] G. Segrè, *Faust in Copenhagen: A struggle for the soul of physics* (Jonathan Cape, London, 2007)
- [10] G. Farmelo, *The Strangest Man. The Hidden Life of Paul Dirac, Quantum Genius* (Faber and Faber, London, 2009)
- [11] D.R. Hartree, *Proc. Camb. Phil. Soc.* **24**, 89,111 (1928)
- [12] J.C. Slater, *Phys. Rev.* **35**, 210 (1930)
- [13] V. Fock, *Z. Physik* **61**, 126 (1930)
- [14] J. C. Slater, *Phys. Rev.* **34**, 1293 (1929)
- [15] F. Bloch, *Z. Physik* **57**, 545 (1929)
- [16] P.A.M. Dirac, *Proc. Roy. Soc. (London) A* **123**, 714 (1929)
- [17] E.H. Lieb, *Rev. Mod. Phys.* **53**, 601 (1981)
- [18] L. Spruch, *Rev. Mod. Phys.* **63**, 151 (1991)
- [19] E. Teller, *Rev. Mod. Phys.* **34**, 627 (1962)
- [20] E.H. Lieb, B. Simon, *Phys. Rev. Lett.* **31**, 681 (1973)
- [21] P. Ehrenfest, *Z. Physik A* **45**, 455 (1927)

- [22] E. Wigner, F. Seitz, Phys. Rev. **43**, 804 (1933); Phys. Rev. **46**, 509 (1934)
- [23] J.C. Slater, Rev. Mod. Phys. **6**, 209 (1934)
- [24] J.C. Slater, Phys. Rev. **81**, 385 (1951)
- [25] J.C. Slater, Phys. Rev. **51**, 846 (1937)
- [26] J. Korrington, Physica **13**, 392 (1947)
- [27] W. Kohn, N. Rostoker, Phys. Rev. **94**, 111 (1954)
- [28] J.C. Slater, Phys. Rev. **165**, 658 (1968)
- [29] R. Gáspár, Acta Phys. Hung. **3**, 263 (1954)
- [30] J.C. Slater, in R. Daudel, B. Pullman (eds.): *The World of Quantum Chemistry* (D. Reidel, Dordrecht, 1974), pp. 3-15
- [31] P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964)
- [32] M. Levy, Proc. Nat. Acad. Sci. (USA), **76**, 6062 (1979)
- [33] See also E.H. Lieb, Int. J. Quantum Chem. **24**, 243 (1983)
- [34] W. Kohn, L.J. Sham, Phys. Rev. **140**, A1133 (1965)
- [35] O. Gunnarsson, B.I. Lundqvist, Phys. Rev. B **13**, 4274 (1976)
- [36] J. Harris, R.O. Jones, J. Phys. F **4**, 1170 (1974)
- [37] V. Heine, in Ψ_k Newsletter, 50, 7-19 (April 2002)
Ab Initio (from Electronic Structure) Calculations of Complex Processes in Materials
http://www.psi-k.org/newsletters/News_50/newsletter_50.pdf
- [38] P.W. Anderson, La Recherche **11**, 98 (1980)
- [39] P. Mavropoulos, private communication
- [40] O. Gunnarsson, J. Harris, R.O. Jones, J. Chem. Phys. **67**, 3970 (1977)
- [41] R.O. Jones, J. Chem. Phys. **71**, 1300 (1979)
- [42] R.O. Jones: *Density Functional Theory: A Personal View*
in A. Avella and F. Mancini (eds.): *Strongly Correlated Systems*
(Springer, Heidelberg, 2012), pp. 1-28
- [43] R.O. Jones, D. Hohl, J. Chem. Phys. **92**, 6710 (1990)
- [44] A.D. Becke, Phys. Rev. A **38**, 3098 (1988)

- [45] B.G. Johnson, P.M.W. Gill, J.A. Pople, *J. Chem. Phys.* **97**, 7846 (1992)
- [46] B.G. Johnson, P.M.W. Gill, J.A. Pople, *J. Chem. Phys.* **98**, 5612 (1993)
- [47] See, for example, R.M. Dreizler, E.K.U. Gross, *Density Functional Theory* (Springer, Heidelberg, 1990)
- [48] N.D. Mermin, *Phys. Rev.* **137**, A1441 (1965)
- [49] U. von Barth, L. Hedin, *J. Phys. C: Solid State Phys.* **5**, 1629 (1972)
- [50] G. Vignale, M. Rasolt, *Phys. Rev. B* **37**, 10685 (1978).
- [51] E. Runge, E.K.U. Gross, *Phys. Rev. Lett.* **52**, 997 (1984)
- [52] O. Gunnarsson, R.O. Jones, *Phys. Rev. B* **31**, 7588 (1985)
- [53] J.P. Perdew, *Phys. Rev. B* **33**, 8822 (1986)
- [54] C.T. Lee, W.T. Yang, R.G. Parr, *Phys. Rev. B* **37**, 785 (1988)
- [55] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985)
- [56] R.O. Jones, *Angew. Chemie* **103**, 647 (1991); *Angew. Chemie Int. Ed.* **30**, 630 (1991)
- [57] See, for example, P. Carloni, U. Röthlisberger, M. Parrinello, *Acc. Chem. Res.* **35**, 455 (2002)
- [58] H. Lin, D.G. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2007)
- [59] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **97**, 146401 (2007)
- [60] S. Kümmel, L. Kronik, *Rev. Mod. Phys.* **80**, 3 (2008)
- [61] B. Kaduk, T. Kowalczyk, T. Van Voorhis, *Chem. Rev.* **112**, 321 (2002)
- [62] X. Ren, P. Rinke, C. Joas, M. Scheffler, *J. Mater. Sci.* **47**, 7447 (2012)
and references therein
- [63] B. Xiao, J.W. Sun, A. Ruzsinszky, J. Feng, J.P. Perdew, *Phys. Rev. B* **86**, 094109 (2013)
- [64] A.D. Becke, *J. Chem. Phys.* **98**, 5648 (1993)
- [65] K. Burke, *J. Chem. Phys.* **136**, 150901 (2012)
- [66] A.D. Boese, N.C. Handy, *J. Chem. Phys.* **114**, 5497 (2001)
- [67] Y. Zhao, D.G. Truhlar, *Acc. Chem. Res.* **41**, 157 (2008), and references therein
- [68] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996)

- [69] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003)
- [70] See, for example, F. Furche, J.P. Perdew, J. Chem. Phys. **124**, 044103 (2006)
- [71] M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B.I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004)
- [72] S. Grimme, J. Comput. Chem. **27**, 1787 (2006)
- [73] A. Tkatchenko, M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009)
- [74] V.I. Anisimov, F. Aryasetiawan, A.I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997)
- [75] M. Cococcioni: *The LDA+U Approach: A Simple Hubbard Correction for Correlated Ground States* in [3]
- [76] M. Cococcioni, S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005), and references therein
- [77] J.J. Shepherd, G.H. Booth, A. Alavi, J. Chem. Phys. **136**, 244101 (2012)
- [78] G.H. Booth, A. Grüneis, G. Kresse, A. Alavi, Nature **453**, 365 (2013)
- [79] L. Delle Site, L.M. Ghiringhelli, D.M. Ceperley, Int. J. Quantum Chem. **113**, 155 (2013)
- [80] J. Kalikka, J. Akola, J. Larrucea, R.O. Jones, Phys. Rev. B **86**, 144113 (2012), and (to be published)
- [81] P.W. Anderson, Ref. [2], pp. 120-130
- [82] B. Civalleri, D. Presti, R. Dovesi, A. Savin, Chem. Modell. **9**, 168 (2012)

2 Many-Electron States

Erik Koch

Computational Materials Science

German Research School for Simulation Sciences

Contents

1	Indistinguishable particles	2
1.1	Symmetric and antisymmetric wave functions	4
2	Reduced density matrices	6
3	Slater determinants	8
3.1	Hartree-Fock	10
4	Second quantization	12
4.1	Creation and annihilation operators	13
4.2	Representation of Slater determinants	15
4.3	Representation of n -body operators	16
4.4	Vacuum state and electron-hole transformation	18
5	Many-body states	20
5.1	Hubbard model	20
5.2	BCS state	23
6	Conclusions	24

1 Indistinguishable particles

Everyday experience tells us that no two objects are the same. We can always find some properties in which they differ. We can even tell identical twins apart, if only we know them well enough: their characteristic traits give them individuality. It has therefore been argued that objects that cannot be distinguished must be identical, as Leibnitz did with his Principle of the Identity of Indiscernibles [1]. We might, however, imagine a replicator that could produce a perfect clone that is identical to the original in every respect. Still, being material objects, original and clone cannot be at the same place at the same time. So even then we could tell original and clone apart by closely following their trajectory in space. This is, of course, only possible in a classical setting. Quantum mechanically, our knowledge of the actual position is limited by the uncertainty principle.

While the idea of identical clones sounds like science fiction, it is standard fare in modern physics: with the discovery of the periodic table it was realized that all materials are built from a small set of different types of atoms, the elementary particles of chemistry. The notion of elementary particle seems, however, to depend on the energy range of interest. While from a chemist's point of view all atoms of a given element are identical, probing the atom at higher energies, we can actually find an internal structure, allowing us to distinguish atoms of the same type but in different excited states [2]. Probing at even higher energies, it turns out that atoms are built of more elementary particles: electrons and the nuclei. These are the elementary particles of condensed-matter physics and quantum chemistry. At still higher energies the nuclei turn out to be built of protons and neutrons, which at even higher energies appear to be built of *up* and *down* quarks.

The elementary particle we will mainly be concerned with here is the electron. For a system of two electrons we can write the wave function as $\Psi(x_1, x_2)$, where x_1 are the degrees of freedom, e.g., position and spin, of the first electron, and x_2 those of the second. As indistinguishable particles, the labeling as first and second electron is of course arbitrary, and we can ask how the wave function changes when we exchange the labels, putting the first electron at x_2 and the second at x_1 . Such a reordering is performed by the permutation operator P :

$$P\Psi(x_1, x_2) = \Psi(x_2, x_1).$$

Indistinguishability implies that the observables do not change under a relabeling of the particles. This is true, in particular, for the probability density: $|\Psi(x_1, x_2)|^2 = |\Psi(x_2, x_1)|^2$, i.e.,

$$P\Psi(x_1, x_2) = e^{i\varphi}\Psi(x_1, x_2) \tag{1}$$

with some phase φ . When permuting twice gives the identity, $P^2 = \mathbb{1}$, then $e^{2i\varphi} = 1$, i.e., φ can only take two different values: $\varphi = 2\pi$, meaning that the wave function does not change (*symmetric*), or $\varphi = \pi$, which means that it changes sign (*antisymmetric*) under the permutation P . These are the irreducible representations of the permutation group. A particular consequence of antisymmetry is that for $\Psi(x_1, x_2 \rightarrow x_1) = 0$, i.e., the two particles can never be found at the same place. This is the Pauli principle.

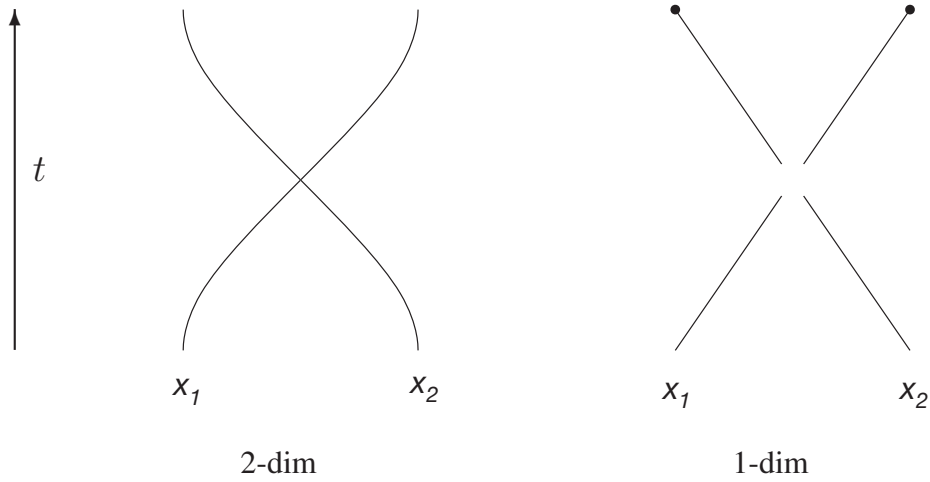


Fig. 1: *Permutation of particles: In 3-dimensional space the permutation is independent of the path along which the particles are exchanged. In 2-dimensional space it matters how the exchange paths wind around each other, giving rise to the braid group and fractional statistics. In 1-dimension, particles have to pass through each other in order to exchange their positions.*

The definition of indistinguishability is that no experiment can distinguish one particle from the other. Consequently, observables involving indistinguishable particles must remain unchanged when the particles are relabeled, or, more technically, they must commute with all possible permutations of the particles. This applies, in particular, to the Hamiltonian: $[P, H] = 0$. This implies that the symmetric and antisymmetric components of the many-body wave function are not mixed by the Hamiltonian: if the initial wave function is symmetric/antisymmetric, this does not change under time evolution.

There is an intriguing connection between the spin of the indistinguishable particles and the symmetry of their many-body wave function: for particles with integer spin (bosons) the wave function is symmetric under particle permutations, for fermions (half-integer spin) the wave function is antisymmetric. In non-relativistic quantum mechanics this *spin-statistics connection* is incorporated *ad hoc* via the initial conditions. In relativistic field-theory the connection between spin and statistics can be derived under fairly general assumptions on the axioms of the theory [3, 4]. For popular accounts, see [5, 6]. More recently there have been efforts to establish the spin-statistics connection in non-relativistic quantum mechanics. The basic idea of the approach is to perform the permutation of particles along a smooth path, where the spin picks up a geometric phase φ [7].

The concept of permuting particles by moving them along paths is also vital for understanding the statistics in lower-dimensional systems. Let us permute two particles by moving particle one along path $\gamma_1(t)$ from $\gamma_1(0) = x_1$ to $\gamma_1(1) = x_2$ and the other particle along $\gamma_2(t)$ from x_2 to x_1 . If we call this operation P_γ , then P_γ^2 is given by moving particle one first along $\gamma_1(t)$ from x_1 to x_2 and then along $\gamma_2(t)$ from x_2 back to x_1 and likewise for the other particle. In three and higher dimensions these combined paths can be continuously deformed into the paths $\iota_1(t) = x_1$ and $\iota_2(t) = x_2$, which correspond to not moving the particles at all, i.e., the identity. Since the paths are homotopic, $P_\gamma^2 = \mathbb{1}$, as assumed above. In two dimensions this is not the

case. Let us assume the two paths $\gamma_1(t)$ and $\gamma_2(t)$ that exchange the particles wind around each other in clockwise direction as shown in Fig. 1. Applying this operation a second time, we obtain paths winding around each other twice to restore the original order of particles. These are however not homotopic to the paths corresponding to the identity $\mathbb{1}$, as deforming $\gamma_1(t)$ and $\gamma_2(t)$ into ι_1 and ι_2 would involve passing the curves through one another. Thus in two dimensions P_γ^2 need not be the identity and thus there is no restriction on the phase φ in (1). Since any phase is allowed, particles with such statistics are called anyons [8]. They appear, e.g., as quasiparticles in the fractional quantum Hall effect.

In one dimension two particles would have to pass through each other to exchange their positions. Therefore particles that cannot be at the same position, as is true for fermions, cannot exchange their positions. Then configuration space splits into equivalent parts, each with a specific ordering of the particles, separated from each other by nodes in the wave function where the coordinates of at least two particles agree. In each of these nodal pockets the ground state wave function is non-vanishing [9]. This is what makes many one-dimensional systems solvable [10].

1.1 Symmetric and antisymmetric wave functions

The (anti)symmetry of a many-body wave function has profound effects on the physical properties of the system. This can already be seen for a simple system of two particles, with one particle in a state $\varphi_a(x)$ and the other in state $\varphi_b(x)$. When the particles are distinguishable the many-body wave function could be

$$\Psi_{12}(x_1, x_2) = \varphi_a(x_1)\varphi_b(x_2) \quad \text{or} \quad \Psi_{21}(x_1, x_2) = \varphi_b(x_1)\varphi_a(x_2). \quad (2)$$

For indistinguishable particles the wave functions is (anti)symmetric

$$\Psi_\pm(x_1, x_2) = \frac{1}{\sqrt{2}}(\Psi_{12}(x_1, x_2) \pm \Psi_{21}(x_1, x_2)). \quad (3)$$

We can then calculate the expectation value of the squared distance

$$\langle (x_1 - x_2)^2 \rangle = \langle x_1^2 \rangle + \langle x_2^2 \rangle - 2 \langle x_1 x_2 \rangle. \quad (4)$$

For wave function Ψ_{12} , assuming that the single-electron states are normalized, we obtain

$$\begin{aligned} \langle x_1^2 \rangle_{12} &= \int dx_1 x_1^2 |\varphi_a(x_1)|^2 \int dx_2 |\varphi_b(x_2)|^2 = \langle x^2 \rangle_a \cdot 1 \\ \langle x_2^2 \rangle_{12} &= \int dx_1 |\varphi_a(x_1)|^2 \int dx_2 x_2^2 |\varphi_b(x_2)|^2 = 1 \cdot \langle x^2 \rangle_b \\ \langle x_1 x_2 \rangle_{12} &= \int dx_1 x_1 |\varphi_a(x_1)|^2 \int dx_2 x_2 |\varphi_b(x_2)|^2 = \langle x \rangle_a \cdot \langle x \rangle_b \end{aligned}$$

Giving the expectation value in terms of single-electron expectation values

$$\langle (x_1 - x_2)^2 \rangle_{12} = \langle x^2 \rangle_a + \langle x^2 \rangle_b - 2 \langle x \rangle_a \langle x \rangle_b. \quad (5)$$

Due to the symmetry $(x_1 - x_2)^2 = (x_2 - x_1)^2$ we obtain the same expectation value for Ψ_{21} . For indistinguishable particles additional cross terms appear in the expectation value of an operator M

$$\langle M \rangle_\pm = \frac{1}{2} \left(\langle M \rangle_{12} \pm \langle \Psi_{12} | M | \Psi_{21} \rangle \pm \langle \Psi_{21} | M | \Psi_{12} \rangle + \langle M \rangle_{21} \right). \quad (6)$$

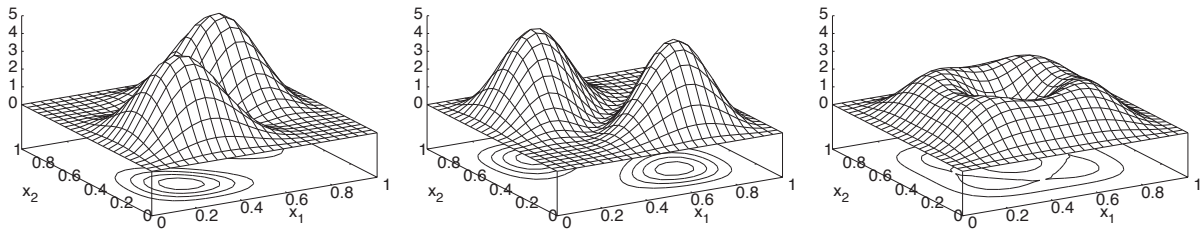


Fig. 2: Probability distribution $|\Psi(x_1, x_2)|^2$ for two identical particles in a one-dimensional infinite potential well with one particle in the ground and the other in the first excited state. For the symmetric wave function, shown on the left, the probability of finding the electrons is largest on the line $x_1 = x_2$, for the antisymmetric wave function, shown in the centre, the probability vanishes there. For comparison, the right-most plot shows the probability distribution for independent particles.

For observables involving only one coordinate like $M = x_1^2$, and similarly for x_2^2 , these terms are of the form

$$\langle \Psi_{12} | x_1^2 | \Psi_{21} \rangle = \int dx_1 x_1^2 \overline{\varphi_a(x_1)} \varphi_b(x_1) \int dx_2 \overline{\varphi_b(x_2)} \varphi_a(x_2), \quad (7)$$

which vanishes if the two states φ_a and φ_b are orthogonal. For operators like $M = x_1 x_2$ involving both coordinates they do not vanish, even for orthogonal states

$$\langle \Psi_{12} | x_1 x_2 | \Psi_{21} \rangle = \int dx_1 x_1 \overline{\varphi_a(x_1)} \varphi_b(x_1) \int dx_2 x_2 \overline{\varphi_b(x_2)} \varphi_a(x_2) = \langle x \rangle_{ab} \cdot \overline{\langle x \rangle_{ab}}. \quad (8)$$

These non-vanishing cross terms are called exchange terms. They make expectation values for symmetric and antisymmetric wave functions different. In the present case

$$\langle (x_1 - x_2)^2 \rangle_{\pm} = \langle x^2 \rangle_a + \langle x^2 \rangle_b - 2 \left(\langle x \rangle_a \langle x \rangle_b \pm |\langle x \rangle_{ab}|^2 \right) \quad (9)$$

we see that the exchange terms decrease (increase) the expectation value of the squared distance by $2|\langle x \rangle_{ab}|^2$ for symmetric (antisymmetric) wave functions compared to the result for distinguishable particles. I.e., indistinguishable fermions tend to avoid each other while bosons tend to move closer together. For two identical particles in a one-dimensional box this tendency is readily apparent from the probability density $|\Psi(x_1, x_2)|^2$ shown in Fig. 2.

The effect of (anti)symmetry thus has to do with the overlap of the of the single-particle states that are involved. When this overlap vanishes for some reason, the symmetry of the wave function makes no difference. An extreme example is two electrons that are strictly localized in non-overlapping regions in space. In this case all integrals of the type (8) vanish, and there is no observable to distinguish an (anti)symmetric from a non-symmetrized state. This makes sense, since their localization in different regions of space makes them actually distinguishable. Such a situation is, of course, never perfectly realized. And in principle we would have to antisymmetrize the states of all electrons in the universe. Except for the rare case that we have sent an electron of an entangled pair to our far-away friend Bob, it is, however, safe to assume

that the electrons on the moon have negligible overlap with those in our (terrestrial) laboratory. We can then consider them distinguishable from those in our experiment, so that we need only antisymmetrize the wave function with respect to the electrons in our apparatus.

Often we have a similar situation where we can use the spin to tell electrons apart: When the Hamiltonian of our system does not affect the spin $[\vec{S}, H] = 0$, and we are only interested in observables that commute with the spin, we can distinguish two types of electrons by their spin direction S_z . In that case we need not antisymmetrize all electrons, but only the spin-up and the spin-down electrons separately. This is a typical example of how quantum numbers that are conserved in the processes we are interested in make elementary particles distinguishable. This is how the concept of elementary particle becomes dependent on what energy scale we are interested in.

2 Reduced density matrices

By definition, observables on an N -particle Hilbert space that do not distinguish between the particles must be symmetric under particle permutations. For example, a single-particle operator $M(x)$ takes the form $\sum_{i=1}^N M(x_i)$ in the N -particle Hilbert space. We can write a general operator as a sum of n -particle operators

$$M(\mathbf{x}) = M_0 + \sum_i M_1(x_i) + \frac{1}{2!} \sum_{i \neq j} M_2(x_i, x_j) + \frac{1}{3!} \sum_{i \neq j \neq k} M_3(x_i, x_j, x_k) + \cdots \quad (10)$$

$$= M_0 + \sum_i M_1(x_i) + \sum_{i < j} M_2(x_i, x_j) + \sum_{i < j < k} M_3(x_i, x_j, x_k) + \cdots, \quad (11)$$

where the summations can be restricted since the operators must be symmetric in their arguments, e.g. $M_2(x_i, x_j) = M_2(x_j, x_i)$, while for two or more identical coordinates the operator is really one of lower order, e.g. $M_2(x_i, x_i)$ only acts on a single coordinate and should be included in M_1 .

To evaluate expectation values it is useful to introduce *density matrices* [11]

$$\Gamma^{(p)}(x'_1, \dots, x'_p; x_1, \dots, x_p) := \binom{N}{p} \int dx_{p+1} \cdots dx_N \overline{\Psi(x'_1, \dots, x'_p, x_{p+1}, \dots, x_N)} \Psi(x_1, \dots, x_p, x_{p+1}, \dots, x_N), \quad (12)$$

where we integrate over all except p coordinates of the normalized N -particle wave function Ψ . When $x = (r, \sigma)$ denotes the coordinate and the spin of the particle, the integral over x means integration over space and summation over spin. The density matrices are obviously related by

$$\Gamma^{(p)}(x'_1, \dots, x'_p; x_1, \dots, x_p) = \frac{p+1}{N-p} \int dx_{p+1} \Gamma^{(p+1)}(x'_1, \dots, x'_p, x_{p+1}; x_1, \dots, x_p, x_{p+1}) \quad (13)$$

They are Hermitean, e.g. $\Gamma^{(2)}(x'_1, x'_2; x_1, x_2) = \overline{\Gamma^{(2)}(x_1, x_2; x'_1, x'_2)}$, and (anti)symmetric in each set of their arguments, e.g. $\Gamma^{(2)}(x'_1, x'_2; x_1, x_2) = -\Gamma^{(2)}(x'_2, x'_1; x_1, x_2)$. The p -body density matrix contains all the information needed for evaluating expectation values of operators up to

order p . The expectation value of a single-electron operator, e.g., the expectation value of the kinetic energy $T = -1/2 \sum_i \Delta_{r_i}$ is obtained from the one-body density matrix as

$$\langle \Psi | T | \Psi \rangle = -\frac{1}{2} \int dx \Delta_r \Gamma^{(1)}(x'; x) \Big|_{x'=x}, \quad (14)$$

where we first keep $x' \neq x$ to make sure that the derivative only operates on the second argument, $x = (r, \sigma)$, but after that set $x' = x$ so both arguments are summed over. For a local operator like the Coulomb potential we can directly work with the diagonal elements of the density matrix. For the interaction of the electrons with a nucleus of charge Z at R this gives

$$\langle V \rangle = -Z \int dx \frac{\Gamma^{(1)}(x; x)}{|r - R|}. \quad (15)$$

Similarly, the Coulomb repulsion between the electrons is given by

$$\langle U \rangle = \int dx dx' \frac{\Gamma^{(2)}(x, x'; x, x')}{|r - r'|}. \quad (16)$$

We see that for calculating the eigenenergies of a many-body Hamiltonian describing a system of N electrons moving around nuclei of charge Z_α at position R_α

$$H = -\frac{1}{2} \sum_i \Delta_i - \sum_{i,\alpha} \frac{Z_\alpha}{|r_i - R_\alpha|} + \sum_{i < j} \frac{1}{|r_i - r_j|} \quad (17)$$

we do not need the full eigenfunction but only the corresponding one-body density matrix and the diagonal elements of the two-body density matrix. It is then tempting to try to calculate the ground state energy of an N -electron system by finding the two-electron density matrix that leads to the lowest energy expectation value. This is known as *Coulson's Challenge* [12]. The approach is, however, not practical since we know no criterion that would tell us what function of four arguments is actually a fermionic density matrix, i.e., one that can be obtained via (12) from an antisymmetric N -electron wave function. For the single-electron density matrix there is such a criterion: for any $\Gamma(x'; x)$ with eigenvalues $\gamma_i \in [0, 1]$ and trace $\text{Tr} \Gamma(x'; x) = N$ there exists a normalized N -electron wave function with single-electron density matrix $\Gamma(x'; x)$.

Since we made sure that the N -electron wave function is normalized, the diagonal elements of the density matrices have straightforward physical interpretations. From the definition (12) we see that the single-electron density matrix gives the electron density $\Gamma^{(1)}(x; x) = n(x)$, while the two-electron density matrix $2\Gamma^{(2)}(x, x'; x, x') = n(x, x')$ gives the conditional electron density, i.e., the electron density at x' , given that one electron is at x . They are normalized accordingly

$$\int dx \Gamma^{(1)}(x; x) = N \quad \text{and} \quad \int dx dx' \Gamma^{(2)}(x, x'; x, x') = \frac{N(N-1)}{2}. \quad (18)$$

The way the two-electron density differs from the simple product of the one-electron densities describes the correlation of the electrons

$$n(x, x') = n(x)n(x') g(x, x'). \quad (19)$$

The factor $g(x, x')$ is called the *pair-correlation function*. Since by the Pauli principle no two electrons can occupy the same state, it vanishes for $x = x'$. From (13) we find

$$n(x) (N - 1) = \int dx' n(x, x') = n(x) \int dx' n(x') g(x, x')$$

which gives the sum rule

$$\int dx' n(x') (g(x, x') - 1) = -1, \quad (20)$$

which implies that the integrand vanishes for $|r - r'| \rightarrow \infty$. In practice $n(x') (g(x, x') - 1)$ is, as a function of x' , quite localized around x . It is called the *exchange-correlation hole*. With this we can write the Coulomb repulsion energy between the electrons as

$$\langle U \rangle = \frac{1}{2} \int dx dx' \frac{n(x)n(x')}{|r - r'|} + \frac{1}{2} \int dx dx' \frac{n(x)n(x') (g(x, x') - 1)}{|r - r'|}, \quad (21)$$

where the first term is the long ranged Coulomb interaction between the uncorrelated charge densities (Hartree energy), while the second term is the interaction of the charge density with its rather localized exchange-correlation hole.

3 Slater determinants

When dealing with indistinguishable particles, we need only consider many-body wave functions that are (anti)symmetric under particle permutations. This can be ensured by explicitly (anti)symmetrizing an arbitrary wave function

$$\mathcal{S}_{\pm} \Psi(x_1, \dots, x_N) := \frac{1}{\sqrt{N!}} \sum_P (\pm 1)^P \Psi(x_{p(1)}, \dots, x_{p(N)}), \quad (22)$$

where $(\pm 1)^P$ is the parity of the permutation P that maps $n \rightarrow p(n)$. Since there are $N!$ different permutations, this can easily become an extremely expensive operation. Since (anti)symmetrization only involves a relabeling of coordinates, in integrals, i.e., matrix elements, we can save some work by observing that in matrix elements only one of the wave functions needs to be properly (anti)symmetrized [11]

$$\int d\mathbf{x} \overline{(\mathcal{S}_{\pm} \Psi_a(\mathbf{x}))} M(\mathbf{x}) (\mathcal{S}_{\pm} \Psi_b(\mathbf{x})) = \sqrt{N!} \int d\mathbf{x} \overline{\Psi_a(\mathbf{x})} M(\mathbf{x}) (\mathcal{S}_{\pm} \Psi_b(\mathbf{x})), \quad (23)$$

where $\mathbf{x} = x_1, \dots, x_N$ and the observable M commutes with particle permutations.

It is remarkable that for products of single-electron states antisymmetrization can be performed very efficiently: it is simply the prescription for calculating a determinant, which can be calculated with $\mathcal{O}(N^3)$ operations. Interestingly, the corresponding operation for bosons, the symmetrized of a product of single-electron states, called the permanent, cannot be performed effi-

ciently. Given a set of spin-orbitals $\varphi_\alpha(x)$ we write the *Slater determinant*

$$\Phi_{\alpha_1 \dots \alpha_N}(\mathbf{x}) := \mathcal{S}_- \varphi_{\alpha_1}(x_1) \cdots \varphi_{\alpha_N}(x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_{\alpha_1}(x_1) & \varphi_{\alpha_2}(x_1) & \cdots & \varphi_{\alpha_N}(x_1) \\ \varphi_{\alpha_1}(x_2) & \varphi_{\alpha_2}(x_2) & \cdots & \varphi_{\alpha_N}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{\alpha_1}(x_N) & \varphi_{\alpha_2}(x_N) & \cdots & \varphi_{\alpha_N}(x_N) \end{vmatrix}. \quad (24)$$

Obviously, replacing the orbitals by linear combinations $\tilde{\varphi}_{\alpha_n}(x) = \sum_{m=1}^N a_{n,m} \varphi_{\alpha_m}(x)$ among themselves produces the same Slater determinant, merely changing the normalization by $\det(A)$, which is non-zero as long as A is invertible.

For $N = 1$ a Slater determinant is simply the one-electron orbital $\varphi_\alpha(x)$; for $N = 2$ it has the familiar form $(\varphi_\alpha(x)\varphi_\beta(x') - \varphi_\beta(x)\varphi_\alpha(x'))/\sqrt{2}$.

Slater determinants are popular electronic wave functions because operations can be calculated efficiently, even for large numbers N of electrons, using standard methods of linear algebra. As an example, using (23), we see that the overlap of two Slater determinants is simply the determinant of the overlap matrix of their single electron orbitals:

$$\int dx_1 \cdots dx_N \overline{\Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N)} \Phi_{\beta_1 \dots \beta_N}(x_1, \dots, x_N) = \det \left(\langle \varphi_{\alpha_n} | \varphi_{\beta_m} \rangle \right). \quad (25)$$

It follows that Slater determinants constructed from a set of orthonormal spin-orbitals $\varphi_\mu(x)$ are normalized – except when they contain an orbital more than once, in which case the determinant, obeying the Pauli principle, vanishes. Likewise, it follows that two Slater determinants $\Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N)$ and $\Phi_{\beta_1 \dots \beta_N}(x_1, \dots, x_N)$ are orthogonal except when they are built from the same set of orbitals, i.e., $\{\alpha_1, \dots, \alpha_N\} = \{\beta_1, \dots, \beta_N\}$. Thus, if we fix some ordering of the orbitals, e.g., $\alpha_1 < \alpha_2 < \cdots < \alpha_N$, the determinants formed from all possible choices of N spin-orbitals from the set of K orthonormal single-electron functions $\varphi_\mu(x)$ forms an orthonormal set in the N -electron Hilbert space. There are $K \cdot (K-1) \cdot (K-2) \cdots (K-(N-1))$ ways of picking N indices out of K . Since we only use one specific ordering of these indices, we still have to divide by $N!$ to obtain the number of such determinants:

$$\frac{K!}{N!(K-N)!} = \binom{K}{N}. \quad (26)$$

They span the antisymmetrized N -particle Hilbert space. Thus, the choice of an orthonormal set of single-electron functions $\{\varphi_\mu(x) | \mu = 1 \dots K\}$ induces an orthonormal basis

$$\left\{ \Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N) \mid \alpha_1 < \alpha_2 < \cdots < \alpha_N \in \{1, \dots, K\} \right\} \quad (27)$$

in the corresponding N -electron space. Given a set of one-electron functions, we can thus, by the variational principle, approach the exact solution of the many-body problem in the corresponding N -electron Hilbert space by including more and more of these determinants. This is called the *configuration interaction* (CI) method. It becomes exact on this space when we include all $\binom{K}{N}$ basis determinants (exact diagonalization or full CI). Even though these calculation very quickly involve unimaginable numbers of determinants – for $N = 25$ electrons in $K = 100$ orbitals the number of basis functions already exceeds 10^{23} – the result is still not exact, as the single electron basis is not complete. This is illustrated in Fig. 3.

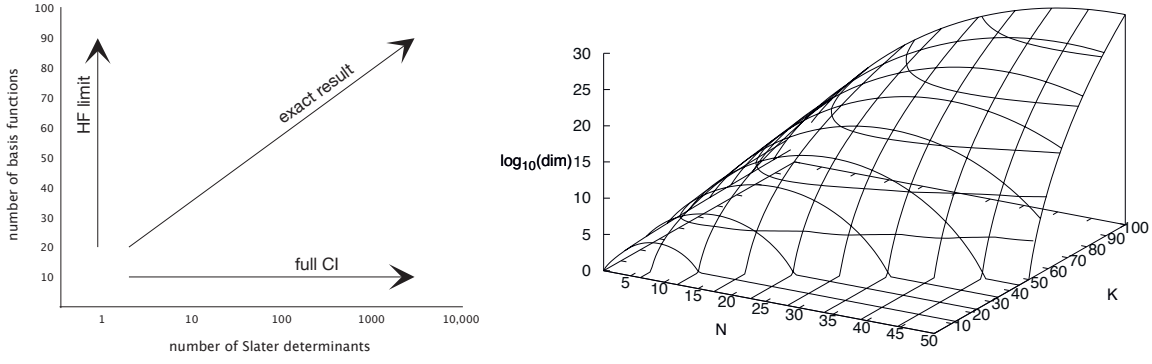


Fig. 3: Left: Convergence of a calculation for an N -electron system with K basis functions. The dimension of the Hilbert space for configuration interaction is $\dim = \binom{K}{N}$. The plot on the right shows the \log_{10} of this as a function of the number of electrons N and orbitals K .

3.1 Hartree-Fock

To calculate expectation values for Slater determinants we take again the route via the reduced density matrices described in Sec. 2. To calculate the one-body density matrix, we expand the Slater determinant along its first row

$$\Phi_{\alpha_1 \dots \alpha_N}(x_1, \dots, x_N) = \frac{1}{\sqrt{N}} \sum_{n=1}^N (-1)^{1+n} \varphi_{\alpha_n}(x_1) \Phi_{\alpha_{i \neq n}}(x_2, \dots, x_N), \quad (28)$$

where $\Phi_{\alpha_{i \neq n}}(x_2, \dots, x_N)$ is the determinant with the first row and the n -th column removed, which can be written as $N-1$ -electron Slater determinants with orbital α_n removed. The integral for obtaining the one-body density matrix is then just of the type (25), so that

$$\Gamma^{(1)}(x'; x) = \frac{1}{N} \sum_{n,m} (-1)^{n+m} \overline{\varphi_{\alpha_n}(x')} \varphi_{\alpha_m}(x) \frac{\det(\langle \varphi_{\alpha_j \neq n} | \varphi_{\alpha_k \neq m} \rangle)}{\det(\langle \varphi_{\alpha_j} | \varphi_{\alpha_k} \rangle)} \quad (29)$$

where we have introduced the normalization factor of the Slater determinant. For orthonormal orbitals this simplifies to the familiar expressions

$$\Gamma^{(1)}(x'; x) = \sum_n \overline{\varphi_{\alpha_n}(x')} \varphi_{\alpha_n}(x) \quad \text{and} \quad n(x) = \sum_n |\varphi_n(x)|^2. \quad (30)$$

For higher-order density matrices, we could expand the $N - 1$ Slater determinants further. A simpler way to generalize (28) is, however, to realize that we can write the permutations of a set of N objects by considering all possible partitions of this set into two sets and taking all permutations among the elements of these sets. This lets us write a Slater determinant as the sum over products of two smaller Slater determinants:

$$\Phi_{\alpha_1 \dots \alpha_N}(\mathbf{x}) = \frac{1}{\sqrt{\binom{N}{p}}} \sum_{n_1 < n_2 < \dots < n_p} (-1)^{1+\sum_i n_i} \Phi_{\alpha_{n_1} \dots \alpha_{n_p}}(x_1, \dots, x_p) \Phi_{\alpha_{i \notin \{n_1, \dots, n_p\}}}(x_{p+1}, \dots, x_N) \quad (31)$$

For $p = 2$ we get the general form of the two-body density matrix for a Slater determinant

$$\Gamma^{(2)}(x'_1, x'_2; x_1, x_2) = \sum_{\substack{n' < m' \\ n < m}} (-1)^{n'+m'+n+m} \overline{\Phi_{\alpha_{n'}, \alpha_{m'}}(x'_1, x'_2)} \Phi_{\alpha_n, \alpha_m}(x_1, x_2) \frac{\det(\langle \varphi_{\alpha_{j \neq n', m'}} | \varphi_{\alpha_{k \neq n, m}} \rangle)}{\det(\langle \varphi_{\alpha_j} | \varphi_{\alpha_k} \rangle)} \quad (32)$$

Since the summation indices are ordered, for orthogonal orbitals only the terms with $(n', m') = (n, m)$ remain, giving the generalization of (30) to $p = 2$

$$\Gamma^{(2)}(x'_1, x'_2; x_1, x_2) = \sum_{n < m} \overline{\Phi_{\alpha_n, \alpha_m}(x'_1, x'_2)} \Phi_{\alpha_n, \alpha_m}(x_1, x_2) \quad (33)$$

and

$$n(x_1, x_2) = \sum_{n, m} |\Phi_{\alpha_n, \alpha_m}(x_1, x_2)|^2, \quad (34)$$

where the factor of 2 is included by summing over all combinations (n, m) , not only the ordered ones, and $m = n$ can be included, since in that case the determinant vanishes. In terms of the orbitals this becomes

$$n(x_1, x_2) = \sum_{n, m} \left(|\varphi_{\alpha_n}(x_1)|^2 |\varphi_{\alpha_m}(x_2)|^2 - \overline{\varphi_{\alpha_n}(x_1)} \varphi_{\alpha_m}(x_1) \overline{\varphi_{\alpha_m}(x_2)} \varphi_{\alpha_n}(x_2) \right), \quad (35)$$

from which it is easy to find the pair correlation function

$$g(x_1, x_2) = 1 - \frac{\sum_{n, m} \overline{\varphi_{\alpha_n}(x_1)} \varphi_{\alpha_m}(x_1) \overline{\varphi_{\alpha_m}(x_2)} \varphi_{\alpha_n}(x_2)}{n(x_1) n(x_2)}. \quad (36)$$

Given the explicit form of the two-body density matrix (32), we can meet Coulson's challenge for the Hamiltonian (17), albeit restricted to density matrices that arise from Slater determinants. This procedure is equivalent to the *Hartree-Fock* method, which gives the Slater determinant for which the total energy is stationary.

For a homogeneous electron gas, i.e., the Hamiltonian (17) without ionic potentials (except for a homogeneous neutralizing background), one such stationary point is, by symmetry, the Slater determinant of plane waves of wave vectors k with $|k| \leq k_F$. For this simple case we can calculate the pair correlation function (36) explicitly

$$g(r_1, \sigma_1, r_2, \sigma_2) - 1 = -9 \frac{(\sin(k_F r) - k_F r \cos(k_F r))^2}{(k_F r)^6} \delta_{\sigma_1, \sigma_2} \quad (37)$$

with $r = r_2 - r_1$. This shows how electrons of the same spin avoid getting close to each other because of the antisymmetry requirement (*exchange hole*), while for a Slater determinant electrons of opposite spin are uncorrelated.

The exchange hole decays rapidly with distance and becomes more localized with increasing density, approaching a delta function in the limit $k_F \rightarrow \infty$. As shown in Fig. 4, the exchange hole is essentially contained in a sphere of the Wigner-Seitz radius $r_\sigma = 2^{1/3} r_s$, i.e., the radius of a sphere containing one electron of a given spin. Since this condition is somewhat

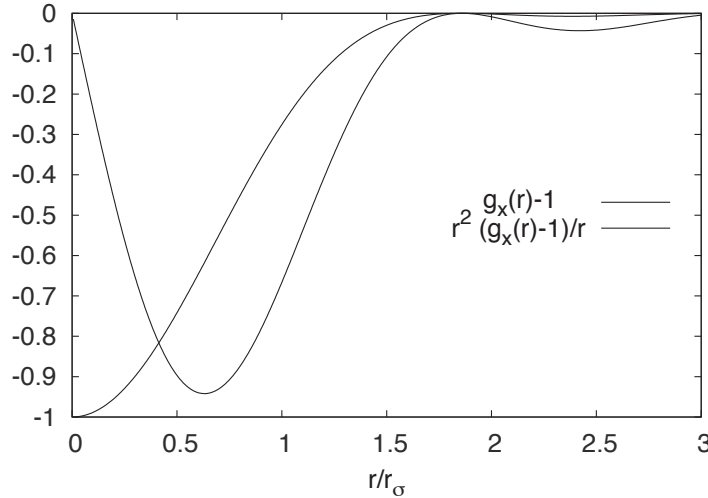


Fig. 4: Exchange hole for a paramagnetic homogeneous electron gas in units of the spin Wigner-Seitz radius $k_F r_\sigma = (9\pi/2)^{1/3}$. In addition, the dotted line shows the contribution of the exchange hole to the Coulomb repulsion energy of Eq. (21).

similar to the sum rule (20) for the pair-correlation function, this is not entirely unexpected. Including correlation effects, which are missing in Hartree-Fock, could increase the range of the exchange-correlation hole, while, missing exchange effects, the correlation hole for electrons of different spin should be more localized. This is in fact what is found in quantum Monte Carlo calculations, see, e.g., Fig. 1 of Ref. [13].

We note in passing that the homogeneous electron gas is not necessarily the Hartree-Fock ground state. Allowing the Slater determinant to break the symmetry of the Hamiltonian, we might obtain a lower energy solution [14]. Enforcing the symmetry of the Slater determinant is called restricted Hartree-Fock, allowing it to have a lower symmetry is unrestricted Hartree-Fock. See, e.g., Ref. [15] for a simple example.

To go beyond Hartree-Fock we could now derive the matrix elements of n -particle operators between different Slater determinants, so that we could represent the operators, e.g., in the orthonormal basis (27). For this we could introduce generalized density matrices with two different many-body wave functions [11]. A much more transparent approach is, however, provided by the formalism of second quantization. It addresses the main inconvenience when working with Slater determinants: keeping track of the sign for sub-determinants. In second quantization these signs are simply stored in the relative positions of certain operators. For this to work, these operators have to change sign when exchanging the order of two of them – they have to anti-commute.

4 Second quantization

The first object to be successfully quantized was the electron. It was no longer described as a classical point-particle but by a quantum mechanical Schrödinger field. Later, for studying the interaction of radiation with matter, also the electromagnetic field had to be quantized, giving rise to quantum particles – photons. This process, pioneered by Dirac [16], was called

the second quantization. Shortly after, Jordan, Klein, and Wigner used a similar approach to quantize the Schrödinger field and found that it could be used to write antisymmetric states in a very convenient way using particle-type operators [17, 18].

When working with Slater determinants of the form (24) we are working in a real-space basis. Like in fundamental quantum mechanics, it is, however, often useful to abstract from a specific basis and work with abstract states: Instead of a wave function $\varphi_\alpha(x)$, we write a Dirac state $|\alpha\rangle$. Second quantization allows us to do the same for Slater determinants.

Let us consider a Slater determinant for two electrons, one in state $\varphi_\alpha(x)$, the other in state $\varphi_\beta(x)$. It is simply the antisymmetrized product of the two states

$$\Phi_{\alpha\beta}(x_1, x_2) = \frac{1}{\sqrt{2}} (\varphi_\alpha(x_1)\varphi_\beta(x_2) - \varphi_\beta(x_1)\varphi_\alpha(x_2)). \quad (38)$$

We could do the same for Dirac states, defining a two-particle Dirac state

$$|\alpha, \beta\rangle := \frac{1}{\sqrt{2}} (|\alpha\rangle|\beta\rangle - |\beta\rangle|\alpha\rangle).$$

The idea of second quantization is then to specify the states using operators

$$c_\beta^\dagger c_\alpha^\dagger |0\rangle = |\alpha, \beta\rangle. \quad (39)$$

When these operators change sign when they are reordered, antisymmetry of the wave function will be automatically ensured

$$|\alpha, \beta\rangle = c_\beta^\dagger c_\alpha^\dagger |0\rangle = -c_\alpha^\dagger c_\beta^\dagger |0\rangle = -|\beta, \alpha\rangle. \quad (40)$$

Naturally, this also implies the Pauli principle for the special case $\beta = \alpha$.

4.1 Creation and annihilation operators

To arrive at the formalism of second quantization we postulate a set of operators that have certain reasonable properties. We then verify that we can use operators with these properties to represent Slater determinants. We start by motivating the properties of the new operators.

To be able to construct many-electron states, we start from the simplest such state: $|0\rangle$ the state with no electron, i.e., the *vacuum state*, which we assume to be normalized $\langle 0|0\rangle = 1$. Next we introduce for each single-electron state $|\alpha\rangle$ (corresponding to an orbital $\varphi_\alpha(x)$) an operator c_α^\dagger . We call it a *creation operator*, since we ask that applying c_α^\dagger to an N -electron state adds an electron in state $|\alpha\rangle$ to that state, making it an $N+1$ electron state. In effect, the operator should be constructed such as to mimic the effect of adding an extra column φ_α and an extra row x_{N+1} to the Slater determinant (24). Since the order in which we add rows/columns matters for the sign of the Slater determinant, we postulate that the operators change sign when exchanged: $c_\alpha^\dagger c_\beta^\dagger = -c_\beta^\dagger c_\alpha^\dagger$. This is more conveniently written as $\{c_\alpha^\dagger, c_\beta^\dagger\} = 0$ by introducing the *anti-commutator*

$$\{A, B\} := AB + BA. \quad (41)$$

The simplest state we can produce with these operators is the single-electron state $|\alpha\rangle = c_\alpha^\dagger|0\rangle$. When we want to calculate its norm, we have to consider the adjoint of $c_\alpha^\dagger|0\rangle$, formally obtaining $\langle\alpha|\alpha\rangle = \langle 0|c_\alpha c_\alpha^\dagger|0\rangle$, or, more generally, $\langle\alpha|\beta\rangle = \langle 0|c_\alpha c_\beta^\dagger|0\rangle$. This must mean that c_α , the adjoint of a creation operator, must remove an electron from the state, otherwise the overlap of $c_\alpha c_\beta^\dagger|0\rangle$ with the vacuum state $\langle 0|$ would vanish. We therefore call the adjoint of the creation operator an *annihilation operator*. We certainly cannot take an electron out of the vacuum state, so $c_\alpha|0\rangle = 0$. Moreover, by taking the adjoint or the anti-commutator of the creation operators, we see that also the annihilation operators anti-commute: $\{c_\alpha, c_\beta\} = 0$. Moreover, to obtain the proper normalization of the single-electron states, we postulate the commutation relation $\{c_\alpha, c_\beta^\dagger\} = \langle\alpha|\beta\rangle$.

Thus, we have defined the vacuum state $|0\rangle$ and the set of operators c_α related to single-electron states $|\alpha\rangle$ with the properties

$$\boxed{\begin{array}{ll} c_\alpha|0\rangle = 0 & \{c_\alpha, c_\beta\} = 0 = \{c_\alpha^\dagger, c_\beta^\dagger\} \\ \langle 0|0\rangle = 1 & \{c_\alpha, c_\beta^\dagger\} = \langle\alpha|\beta\rangle \end{array}} \quad (42)$$

We note that the creators and annihilators are not ordinary operators in a Hilbert space, but transfer states from an N -electron to a $N \pm 1$ -electron Hilbert space, i.e., they are operators defined on the *Fock space*. It is also remarkable that the mixed anti-commutator is the only place where the orbitals that distinguish different operators enter.

One type of operators is particularly useful for making contact with the real-space picture: The operators $\hat{\Psi}^\dagger(x)$, with $x = (r, \sigma)$, that create an electron of spin σ at position r , i.e., in state $|x\rangle = |r, \sigma\rangle$. Because of their importance they get a special name, *field operators*, and a special symbol $\hat{\Psi}^\dagger(x)$ instead of c_x^\dagger , but really they are just ordinary creation operators for the states corresponding to a delta function at r and a spin σ . The anti-commutator for the field-operators obviously follow from (42)

$$\{\hat{\Psi}(x), \hat{\Psi}(x')\} = 0 = \{\hat{\Psi}^\dagger(x), \hat{\Psi}^\dagger(x')\} \quad \text{and} \quad \{\hat{\Psi}(x), \hat{\Psi}^\dagger(x')\} = \delta(x - x'). \quad (43)$$

Given the single-electron wave functions in real space $\varphi_\alpha(x)$, we can express any creation operator in terms of the field operators

$$c_\alpha^\dagger = \int dx \varphi_\alpha(x) \hat{\Psi}^\dagger(x). \quad (44)$$

Using (43), it is easy to see that these operator indeed fulfill all properties (42) required of the creation operators.

Conversely, if we have a complete set of single electron states $\{\varphi_{\alpha_n}(x)\}$, we can expand the field operators in terms of the corresponding creators and annihilators. Given the overlap matrix $S = (\langle\alpha_n|\alpha_m\rangle)$ we can use the Cholesky factorization $S^{-1} = T^\dagger T$ to orthonormalize the orbitals $\tilde{\varphi}_{\alpha_n}(x) = \sum T_{n,m} \varphi_{\alpha_m}(x)$. The completeness relation is then

$$\sum_{n,m} \overline{\varphi_{\alpha_n}(x)} (S^{-1})_{n,m} \varphi_{\alpha_m}(x') = \sum_j \overline{\tilde{\varphi}_{\alpha_j}(x)} \tilde{\varphi}_{\alpha_j}(x') = \delta(x - x'). \quad (45)$$

Using this together with the commutation relations (42) we see that the operators

$$\hat{\Psi}(x) = \sum_n \tilde{\varphi}_{\alpha_n}(x) c_{\alpha_n}, \quad (46)$$

fulfill the commutation relations (43) of the field operators.

4.2 Representation of Slater determinants

We now show that we can write a Slater determinant in terms of the algebra (42) we have just defined. For this we consider an N -electron state $\prod c_{\alpha}^{\dagger} |0\rangle$ and prove that its real-space representation, obtained via the field operators is just the corresponding Slater determinant

$$\Phi_{\alpha_1 \alpha_2 \dots \alpha_N}(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{N!}} \langle 0 | \hat{\Psi}(x_1) \hat{\Psi}(x_2) \dots \hat{\Psi}(x_N) c_{\alpha_N}^{\dagger} \dots c_{\alpha_2}^{\dagger} c_{\alpha_1}^{\dagger} | 0 \rangle \quad (47)$$

Not surprisingly, the proof is by induction. As a warm-up we consider the case of a single-electron wave function ($N = 1$). Using the special case of an anti-commutation relation

$$\{\hat{\Psi}(x), c_{\alpha}^{\dagger}\} = \int dx' \varphi_{\alpha}(x') \{\hat{\Psi}(x), \hat{\Psi}^{\dagger}(x')\} = \varphi_{\alpha}(x) \quad (48)$$

we see that

$$\langle 0 | \hat{\Psi}(x_1) c_{\alpha_1}^{\dagger} | 0 \rangle = \langle 0 | \varphi_{\alpha_1}(x_1) - c_{\alpha_1}^{\dagger} \hat{\Psi}(x_1) | 0 \rangle = \varphi_{\alpha_1}(x_1) \quad (49)$$

For the two-electron state $N = 2$, we anticommute $\hat{\Psi}(x_2)$ in two steps to the right

$$\begin{aligned} \langle 0 | \hat{\Psi}(x_1) \hat{\Psi}(x_2) c_{\alpha_2}^{\dagger} c_{\alpha_1}^{\dagger} | 0 \rangle &= \langle 0 | \hat{\Psi}(x_1) (\varphi_{\alpha_2}(x_2) - c_{\alpha_2}^{\dagger} \hat{\Psi}(x_2)) c_{\alpha_1}^{\dagger} | 0 \rangle \\ &= \langle 0 | \hat{\Psi}(x_1) c_{\alpha_1}^{\dagger} | 0 \rangle \varphi_{\alpha_2}(x_2) - \langle 0 | \hat{\Psi}(x_1) c_{\alpha_2}^{\dagger} \hat{\Psi}(x_2) c_{\alpha_1}^{\dagger} | 0 \rangle \\ &= \varphi_{\alpha_1}(x_1) \varphi_{\alpha_2}(x_2) - \varphi_{\alpha_2}(x_1) \varphi_{\alpha_1}(x_2). \end{aligned} \quad (50)$$

We see how anti-commuting automatically produces appropriate sign for the antisymmetric wave function. Dividing by $\sqrt{2}$, we obtain the desired two-electron Slater determinant.

The general case of an N -electron state works just the same. Anti-commuting $\hat{\Psi}(x_N)$ all the way to the right produces $N - 1$ terms with alternating sign

$$\begin{aligned} &\langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) \hat{\Psi}(x_N) c_{\alpha_N}^{\dagger} c_{\alpha_{N-1}}^{\dagger} \dots c_{\alpha_1}^{\dagger} | 0 \rangle = \\ &+ \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) c_{\alpha_{N-1}}^{\dagger} \dots c_{\alpha_1}^{\dagger} | 0 \rangle \varphi_{\alpha_N}(x_N) \\ &- \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) \prod_{n \neq N-1} c_{\alpha_n}^{\dagger} | 0 \rangle \varphi_{\alpha_{N-1}}(x_N) \\ &\vdots \\ &(-1)^N \langle 0 | \hat{\Psi}(x_1) \dots \hat{\Psi}(x_{N-1}) c_{\alpha_N}^{\dagger} \dots c_{\alpha_2}^{\dagger} | 0 \rangle \varphi_{\alpha_1}(x_N) \end{aligned}$$

Using (47) for the $N - 1$ -electron states, this is nothing but the Laplace expansion of

$$D = \begin{vmatrix} \varphi_{\alpha_1}(x_1) & \varphi_{\alpha_2}(x_1) & \cdots & \varphi_{\alpha_N}(x_1) \\ \varphi_{\alpha_1}(x_2) & \varphi_{\alpha_2}(x_2) & \cdots & \varphi_{\alpha_N}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{\alpha_1}(x_N) & \varphi_{\alpha_2}(x_N) & \cdots & \varphi_{\alpha_N}(x_N) \end{vmatrix}$$

along the N th row. Dividing by $\sqrt{N!}$ we see that we have shown (47) for N -electron states. Thus we see that instead of working with the Slater determinant $\Phi_{\alpha_1\alpha_2\ldots\alpha_N}(x_1, x_2, \ldots, x_N)$ we can work with the corresponding N -electron product state $\prod c_{\alpha}^{\dagger} |0\rangle$. In particular, instead of working with the basis of Slater determinants (27) induced by an orthonormal set of single-electron states $\{\varphi_{\alpha_n}(x)\}$, we can work with the corresponding basis of product states

$$\left\{ \prod_{\alpha_1 < \cdots < \alpha_N} c_{\alpha_N}^{\dagger} \cdots c_{\alpha_1}^{\dagger} |0\rangle \right\}. \quad (51)$$

4.3 Representation of n -body operators

Having established the relation between product states and Slater determinants, it is straightforward to express the matrix elements of a general n -body operator (11)

$$M(\mathbf{x}) = M_0 + \sum_i M_1(x_i) + \sum_{i < j} M_2(x_i, x_j) + \sum_{i < j < k} M_3(x_i, x_j, x_k) + \cdots \quad (52)$$

with N -electron Slater determinants:

$$\begin{aligned} \int dx_1 \cdots dx_N \overline{\Phi_{\beta_1 \cdots \beta_N}(x_1, \cdots, x_N)} M(x_1, \cdots, x_N) \Phi_{\alpha_1 \cdots \alpha_N}(x_1, \cdots, x_N) \\ = \left\langle 0 \left| c_{\beta_1} \cdots c_{\beta_N} \hat{M} c_{\alpha_N}^{\dagger} \cdots c_{\alpha_1}^{\dagger} \right| 0 \right\rangle \end{aligned} \quad (53)$$

with the representation of the n -body operator in terms of field operators

$$\hat{M} = \frac{1}{N!} \int dx_1 \cdots dx_N \hat{\Psi}^{\dagger}(x_N) \cdots \hat{\Psi}^{\dagger}(x_1) M(x_1, \cdots, x_N) \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N). \quad (54)$$

Note that this form of the operator is only valid when applied to N -electron states. But from here on, we can work entirely in terms of our algebra (42).

To see what (54) means we look at its parts (52). As usual, we start with the simplest case, the zero-body operator, which, up to trivial prefactor, is $M_0(x_1, \cdots, x_N) = 1$. Operating on an N -electron wave function, it gives

$$\begin{aligned} \hat{M}_0 &= \frac{1}{N!} \int dx_1 dx_2 \cdots dx_N \hat{\Psi}^{\dagger}(x_N) \cdots \hat{\Psi}^{\dagger}(x_2) \hat{\Psi}^{\dagger}(x_1) \hat{\Psi}(x_1) \hat{\Psi}(x_2) \cdots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \int dx_2 \cdots dx_N \hat{\Psi}^{\dagger}(x_N) \cdots \hat{\Psi}^{\dagger}(x_2) \quad \hat{N} \quad \hat{\Psi}(x_2) \cdots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \int dx_2 \cdots dx_N \hat{\Psi}^{\dagger}(x_N) \cdots \hat{\Psi}^{\dagger}(x_2) \quad 1 \quad \hat{\Psi}(x_2) \cdots \hat{\Psi}(x_N) \\ &\vdots \\ &= \frac{1}{N!} 1 \cdot 2 \cdots N = 1 \end{aligned} \quad (55)$$

where we have used that

$$\int dx \hat{\Psi}^{\dagger}(x) \hat{\Psi}(x) = \hat{N} \quad (56)$$

is the number operator and that applying n annihilation operators $\hat{\Psi}(x_j)$ to an N -electron state gives a state with $N - n$ electrons. We note that we obtain a form of $\hat{M}_0 = 1$ that apparently does not depend on the number of electrons in the wave function that it is applied to. This was not the case for the original expression (54).

Next we consider one-body operators $M(x_1, \dots, x_N) = \sum_j M_1(x_j)$

$$\begin{aligned}\hat{M}_1 &= \frac{1}{N!} \int dx_1 \cdots dx_N \hat{\Psi}^\dagger(x_N) \cdots \hat{\Psi}^\dagger(x_1) \sum_j M_1(x_j) \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \sum_j \int dx_j \hat{\Psi}^\dagger(x_j) M_1(x_j) (N-1)! \hat{\Psi}(x_j) \\ &= \frac{1}{N} \sum_j \int dx_j \hat{\Psi}^\dagger(x_j) M_1(x_j) \hat{\Psi}(x_j) \\ &= \int dx \hat{\Psi}^\dagger(x) M_1(x) \hat{\Psi}(x)\end{aligned}$$

Here we have first anticommutated $\hat{\Psi}^\dagger(x_j)$ all the way to the left and $\hat{\Psi}(x_j)$ to the right. Since these take the same numbers of anticommutations, there is no sign involved. The operation leaves the integrals over the variables except x_i , a zero-body operator for $N - 1$ electron states, operating on $\hat{\Psi}(x_j)|N\text{-electron state}\rangle$.

Expanding the field-operators in a complete orthonormal set $\hat{\Psi}(x) = \sum_n \varphi_{\alpha_n}(x) c_{\alpha_n}$ gives

$$\hat{M}_1 = \sum_{n,m} \int dx \overline{\varphi_{\alpha_n}(x)} M(x) \varphi_{\alpha_m}(x) c_{\alpha_n}^\dagger c_{\alpha_m} = \sum_{n,m} \langle \alpha_n | M_1 | \alpha_m \rangle c_{\alpha_n}^\dagger c_{\alpha_m}. \quad (57)$$

Also here we find a form for \hat{M}_1 that is apparently independent of the number of electrons N and can be evaluated directly in the basis states (51).

For the two-body operators $M(x_1, \dots, x_N) = \sum_{i < j} M_2(x_i, x_j)$ we proceed in the familiar way, anti-commuting first the operators with the coordinates involved in M_2 all the way to the left and right. This time we are left with a zero-body operator for $N - 2$ electrons:

$$\begin{aligned}\hat{M}_2 &= \frac{1}{N!} \int dx_1 \cdots dx_N \hat{\Psi}^\dagger(x_N) \cdots \hat{\Psi}^\dagger(x_1) \sum_{i < j} M_2(x_i, x_j) \hat{\Psi}(x_1) \cdots \hat{\Psi}(x_N) \\ &= \frac{1}{N!} \sum_{i < j} \int dx_i dx_j \hat{\Psi}^\dagger(x_j) \hat{\Psi}^\dagger(x_i) M_2(x_i, x_j) (N-2)! \hat{\Psi}(x_i) \hat{\Psi}(x_j) \\ &= \frac{1}{N(N-1)} \sum_{i < j} \int dx_i dx_j \hat{\Psi}^\dagger(x_j) \hat{\Psi}^\dagger(x_i) M_2(x_i, x_j) \hat{\Psi}(x_i) \hat{\Psi}(x_j) \\ &= \frac{1}{2} \int dx dx' \hat{\Psi}^\dagger(x') \hat{\Psi}^\dagger(x) M_2(x, x') \hat{\Psi}(x) \hat{\Psi}(x')\end{aligned}$$

Expanding in an orthonormal basis, we get

$$\begin{aligned}\hat{M}_2 &= \frac{1}{2} \sum_{n,n',m,m'} \int dx dx' \overline{\varphi_{\alpha_{n'}}(x')} \varphi_{\alpha_n}(x) M_2(x, x') \varphi_{\alpha_m}(x) \varphi_{\alpha_{m'}}(x') c_{\alpha_{n'}}^\dagger c_{\alpha_n}^\dagger c_{\alpha_m} c_{\alpha_{m'}} \\ &= \frac{1}{2} \sum_{n,n',m,m'} \langle \alpha_n \alpha_{n'} | M_2 | \alpha_m \alpha_{m'} \rangle c_{\alpha_{n'}}^\dagger c_{\alpha_n}^\dagger c_{\alpha_m} c_{\alpha_{m'}} \quad (58)\end{aligned}$$

where the exchange of the indices in the second line is a consequence of the way the Dirac state for two electrons is usually written: first index for the first coordinate, second index for the second, while taking the adjoint of the operators changes their order. Obviously, from the symmetry $M_2(x, x') = M_2(x', x)$ follows $\langle \alpha_n \alpha_{n'} | M_2 | \alpha_m \alpha_{m'} \rangle = \langle \alpha_{n'} \alpha_n | M_2 | \alpha_{m'} \alpha_m \rangle$.

The procedure generalizes to operators acting on more than two electrons in the natural way.

We note that, while we started from a form of the operators (52) that was explicitly formulated in an N -electron Hilbert space, the results (55), (57), and (58) are of the same form no matter what value N takes. Thus these operators are valid not just on some N -electron Hilbert space, but on the entire Fock space. This is a particular strength of the second quantized formulation.

4.4 Vacuum state and electron-hole transformation

We have introduced the state $|0\rangle$ as the state with no electrons, $N = 0$. The whole formalism of second quantization requires, however, only (42), i.e., that $|0\rangle$ is normalized and annihilated by the annihilation operators. We can exploit this to obtain more convenient descriptions of many-electron systems. As a first example, see [19] for the physics background, let us consider the d -states of an atom. Denoting the operator for putting an electron of spin σ in the d -orbital with directional quantum number m by $d_{m\sigma}^\dagger$, we can describe a d^N configuration, i.e., a state with N d -electrons as a linear combination of product states $\prod_{n=1}^N d_{m_n\sigma}^\dagger |0\rangle$. Here $|0\rangle$ is the state without electrons. This is the description we have used so far. It specifies the states the electrons are in. For an almost full shell it might, however, be more convenient to specify the state in terms of the non-occupied states. We can do this by introducing a new “vacuum” state

$$|\text{full shell}\rangle = d_{-2\downarrow}^\dagger d_{-1\downarrow}^\dagger \cdots d_{2\downarrow}^\dagger d_{-2\uparrow}^\dagger d_{-1\uparrow}^\dagger \cdots d_{2\uparrow}^\dagger |0\rangle = \prod_{\sigma} \prod_{m=2}^{-2} d_{m\sigma}^\dagger |0\rangle, \quad (59)$$

corresponding to a filled d -shell. $|\text{full shell}\rangle$ certainly does not fulfill the requirements for a vacuum state, since $d_{m\sigma} |\text{full shell}\rangle \neq 0$. Thanks to the Pauli principle it is, however, annihilated by any electron creation operator c_δ^\dagger in the space of d -orbitals. Thus, when we relabel these electron creation operators as hole annihilation operators, $h_\delta = c_\delta^\dagger$, then $|\text{full shell}\rangle$ behaves as a vacuum state for these newly labeled operators h_δ . We pick the relation between the hole state δ and the corresponding electron states $\bar{\delta}$ such that form of the anti-commutation relations remain unchanged: $\{h_\alpha, h_\beta^\dagger\} = \{c_\alpha^\dagger, c_\beta\} = \langle \bar{\beta} | \bar{\alpha} \rangle = \langle \alpha | \beta \rangle$. Having established an isomorphism between the algebra of electron operators and that of the corresponding hole operators, we can relate electron expectation values to those of hole-states, e.g., $\langle 0 | c_\alpha c_\beta^\dagger | 0 \rangle = \langle \text{full} | h_{\bar{\alpha}} h_{\bar{\beta}}^\dagger | \text{full} \rangle$.

A common choice is to take the complex conjugate state $\varphi_{\bar{\delta}}(x) = \overline{\varphi_\delta(x)}$.

We can now ask what kind of particles the operators h_δ^\dagger create. This is most easily done in the basis $d_{m\sigma}^\dagger$ of spherical harmonics; the general h_δ^\dagger follow then by expanding them in the $d_{m\sigma}^\dagger$. A full d -shell has total orbital momentum $L = 0$ and total spin $S = 0$. Removing an electron in state $|m\sigma\rangle$ thus changes L_z from 0 to $-m$ and S_z from 0 to $-\sigma$. The corresponding creator therefore creates a hole with directional quantum number $-m$ and spin $-\sigma$. We express this by writing the electron-hole transformation as $h_{m\sigma}^\dagger = d_{-m, -\sigma}$. We can make a similar argument

for a completely filled band

$$|\text{full band}\rangle = \prod_{\sigma} \prod_k b_{k\sigma}^{\dagger} |0\rangle, \quad (60)$$

with hole operators $h_{k\sigma}^{\dagger} = b_{-k,-\sigma}$.

We can then relate the matrix elements for N -electron states of type $|e\rangle = \prod c_{\alpha_n}^{\dagger} |0\rangle$ and the related N -hole states $|h\rangle = \prod h_{\bar{\alpha}_n}^{\dagger} |\text{full}\rangle$, where $|\text{full}\rangle = \prod_{n=1}^{N_{\text{states}}} c_{\alpha_n}^{\dagger} |0\rangle$ is assumed to be normalized, as is required of a vacuum state. Working with orthonormal operators, we find that the matrix elements for a one-body operator (57) change sign and have a constant shift on the diagonal

$$\langle h' | \hat{M}_1 | h \rangle = \sum_{n,m} \langle \alpha_n | M_1 | \alpha_m \rangle \langle h' | c_{\alpha_n}^{\dagger} c_{\alpha_m} | h \rangle \quad (61)$$

$$= \sum_{n,m} \langle \alpha_n | M_1 | \alpha_m \rangle \langle h' | h_{\bar{\alpha}_n} h_{\bar{\alpha}_m}^{\dagger} | h \rangle \quad (62)$$

$$= \sum_{n,m} \langle \alpha_n | M_1 | \alpha_m \rangle \left(\underbrace{\langle \bar{\alpha}_n | \bar{\alpha}_m \rangle}_{\langle \text{full} | c_{\alpha_n}^{\dagger} c_{\alpha_m} | \text{full} \rangle} \langle h' | h \rangle - \underbrace{\langle h' | h_{\bar{\alpha}_m}^{\dagger} h_{\bar{\alpha}_n} | h \rangle}_{=\langle e' | c_{\alpha_m}^{\dagger} c_{\alpha_n} | e \rangle} \right) \quad (63)$$

$$= \langle \text{full} | \hat{M}_1 | \text{full} \rangle \delta_{h',h} - \langle e' | \hat{M}_1 | e \rangle. \quad (64)$$

In going to the second line, we converted from writing the matrix element in electron operators to the formulation in hole operators. The identity of the matrix elements for the N -hole and N -electron states in the third line follows from the fact that the operators $c_{\bar{\alpha}}$ and h_{α} form, with their respective vacua, the same algebra. For two-body operators (58) we use

$$h_{\alpha} h_{\beta} h_{\gamma}^{\dagger} h_{\delta}^{\dagger} = h_{\delta}^{\dagger} h_{\gamma}^{\dagger} h_{\beta} h_{\alpha} - \langle \alpha | \gamma \rangle h_{\beta} h_{\delta}^{\dagger} + \langle \alpha | \delta \rangle h_{\beta} h_{\gamma}^{\dagger} - \langle \beta | \gamma \rangle h_{\delta}^{\dagger} h_{\alpha} + \langle \beta | \delta \rangle h_{\gamma}^{\dagger} h_{\alpha}. \quad (65)$$

Collecting contributions of the direct two-body terms to $\langle h' | \hat{M}_2 | h \rangle$ we get

$$\frac{1}{2} \sum_{\alpha\beta\gamma\delta} \langle \beta\alpha | M_2 | \gamma\delta \rangle \langle h' | \delta_{\alpha\delta} h_{\beta} h_{\gamma}^{\dagger} - \delta_{\beta\gamma} h_{\delta}^{\dagger} h_{\alpha} | h \rangle = \frac{1}{2} \sum_{\alpha\beta} \langle \beta\alpha | M_2 | \beta\alpha \rangle \delta_{h',h} - \sum_{\alpha\beta\gamma} \langle \beta\alpha | \gamma\alpha \rangle \langle h' | h_{\gamma}^{\dagger} h_{\beta} | h \rangle$$

and similarly for minus the exchange terms

$$\frac{1}{2} \sum_{\alpha\beta\gamma\delta} \langle \beta\alpha | M_2 | \gamma\delta \rangle \langle h' | \delta_{\alpha\gamma} h_{\beta} h_{\delta}^{\dagger} - \delta_{\beta\delta} h_{\gamma}^{\dagger} h_{\alpha} | h \rangle = \frac{1}{2} \sum_{\alpha\beta} \langle \beta\alpha | M_2 | \alpha\beta \rangle \delta_{h',h} - \sum_{\alpha\beta\gamma} \langle \beta\alpha | \alpha\gamma \rangle \langle h' | h_{\gamma}^{\dagger} h_{\beta} | h \rangle$$

The first terms only contribute to diagonal matrix elements and give the expectation value of the full shell $\langle \text{full} | \hat{M}_2 | \text{full} \rangle$. The one-body terms also contribute only to the diagonal when the full shell is symmetric (atomic shell: radial symmetry, filled band: $k = 0$) and M_2 conserves the corresponding quantum numbers (atomic shell: $m_1 + m_2 = m_3 + m_4$, filled band: $k_1 + k_2 = k_3 + k_4$): fixing, e.g., $\alpha = \gamma$ then also fixes $\beta = \delta$. Moreover, all terms $\sum_{\alpha} \langle \beta\alpha | M_2 | \beta\alpha \rangle$ or the corresponding exchange term are independent of $|\beta\rangle$ for orbitals of the same symmetry (just rotate the basis to the desired $|\beta'\rangle$) so that, again, there is just a constant shift of the diagonal elements

$$\langle h' | \hat{M}_2 | h \rangle = \left(\langle \text{full} | \hat{M}_2 | \text{full} \rangle + N \sum_{\alpha} (\langle \beta\alpha | M_2 | \beta\alpha \rangle - \langle \beta\alpha | M_2 | \alpha\beta \rangle) \right) \delta_{e',e} + \langle e' | \hat{M}_2 | e \rangle. \quad (66)$$

An interesting new situation arises when we consider product states that are not closed shells. A popular example is the Fermi sea for a homogeneous electron gas

$$|\text{Fermi sea}\rangle = \prod_{\sigma} \prod_{|k| \leq k_F} c_{k\sigma}^{\dagger} |0\rangle. \quad (67)$$

We can now introduce new annihilation operators as

$$h_{k\sigma} = \begin{cases} c_{-k,-\sigma}^{\dagger} & \text{for } |k| \leq k_F \\ c_{k,\sigma} & \text{for } |k| > k_F \end{cases} \quad (68)$$

They are of hole-type for states occupied in $|\text{Fermi sea}\rangle$, while for empty states they are of electron type. This mixing of character has an interesting consequence: electron creation operators in a basis other than that used for defining the new vacuum are transformed to operators with mixed creator/annihilator contributions. As an example, the field operator

$$\hat{\Psi}_{\sigma}(r) = \int dk e^{ikr} c_k = \int_{|k| \leq k_F} dk e^{ikr} h_{-k,-\sigma}^{\dagger} + \int_{|k| > k_F} dk e^{ikr} h_{k,\sigma} \quad (69)$$

is no longer a pure annihilation operator in the hole picture. I.e., we no longer get the full algebra (42) but are restricted to operators defined in the basis that was used to generate the new vacuum.

5 Many-body states

We now consider small model Hamiltonians to illustrate the techniques introduced so far. This will also allow us to discuss characteristic many-body states without too much complication.

5.1 Hubbard model

As the first example we study the Hubbard model with two sites, $i = 1, 2$, between which the electrons can hop with matrix element $-t$ and with an on-site Coulomb repulsion U

$$H = -t \sum_{\sigma} \left(c_{2\sigma}^{\dagger} c_{1\sigma} + c_{1\sigma}^{\dagger} c_{2\sigma} \right) + U \sum_{i \in \{1,2\}} n_{i\uparrow} n_{i\downarrow}. \quad (70)$$

The number of electrons N and S_z are conserved, so the Fock space Hamiltonian is block-diagonal in the Hilbert spaces with fixed number of up- and down-spin electrons N_{\uparrow} and N_{\downarrow} with dimensions

N	0	1	2	3	4	
N_{\uparrow}	0	1 0	2 1 0	2 1 2	2	
N_{\downarrow}	0	0 1	0 1 2	1 2 2	2	
dim	1	2 2	1 4 1	2 2 1	1	16

The Hamiltonian for $N = N_\uparrow = 1$ is easily constructed. By introducing the basis states $c_{1\uparrow}^\dagger|0\rangle$ and $c_{2\uparrow}^\dagger|0\rangle$, we obtain the Hamiltonian matrix

$$\langle 0 | \begin{pmatrix} c_{1\uparrow} \\ c_{2\uparrow} \end{pmatrix} H \begin{pmatrix} c_{1\uparrow}^\dagger & c_{2\uparrow}^\dagger \end{pmatrix} | 0 \rangle = \begin{pmatrix} 0 & -t \langle 0 | c_{1\uparrow} c_{1\uparrow}^\dagger c_{2\uparrow} c_{2\uparrow}^\dagger | 0 \rangle \\ -t \langle 0 | c_{2\uparrow} c_{2\uparrow}^\dagger c_{1\uparrow} c_{1\uparrow}^\dagger | 0 \rangle & 0 \end{pmatrix} = \begin{pmatrix} 0 & -t \\ -t & 0 \end{pmatrix}.$$

This is easily diagonalized giving the familiar bonding and antibonding solution

$$|\pm\rangle = \frac{1}{\sqrt{2}} \left(c_{1\uparrow}^\dagger \pm c_{2\uparrow}^\dagger \right) | 0 \rangle = c_{\pm\uparrow}^\dagger | 0 \rangle. \quad (71)$$

For $N_\uparrow = 1 = N_\downarrow$, we obtain a non-trivial interacting system

$$\langle 0 | \begin{pmatrix} c_{1\uparrow}c_{2\downarrow} \\ c_{2\uparrow}c_{1\downarrow} \\ c_{1\uparrow}c_{1\downarrow} \\ c_{2\uparrow}c_{2\downarrow} \end{pmatrix} H \begin{pmatrix} c_{2\downarrow}^\dagger c_{1\uparrow}^\dagger & c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger & c_{1\downarrow}^\dagger c_{1\uparrow}^\dagger & c_{2\downarrow}^\dagger c_{2\uparrow}^\dagger \end{pmatrix} | 0 \rangle = \begin{pmatrix} 0 & 0 & -t & -t \\ 0 & 0 & -t & -t \\ -t & -t & U & 0 \\ -t & -t & 0 & U \end{pmatrix}. \quad (72)$$

To diagonalize the matrix, we transform the basis into linear combinations of covalent and ionic states

$$|\text{cov}_\pm\rangle = \frac{1}{\sqrt{2}} \left(c_{2\downarrow}^\dagger c_{1\uparrow}^\dagger \pm c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger \right) | 0 \rangle \quad (73)$$

$$|\text{ion}_\pm\rangle = \frac{1}{\sqrt{2}} \left(c_{1\downarrow}^\dagger c_{1\uparrow}^\dagger \pm c_{2\downarrow}^\dagger c_{2\uparrow}^\dagger \right) | 0 \rangle \quad (74)$$

It is then easy to verify that $|\text{cov}_-\rangle$ is an eigenstate with eigenvalue $\varepsilon_{\text{cov}_-} = 0$ and that $|\text{ion}_-\rangle$ has eigenenergy $\varepsilon_{\text{ion}_-} = U$. The remaining two states mix

$$\begin{pmatrix} \langle \text{cov}_+ | \\ \langle \text{ion}_+ | \end{pmatrix} H \begin{pmatrix} |\text{cov}_+\rangle & |\text{ion}_+\rangle \end{pmatrix} = \frac{1}{2} \left\{ U - \begin{pmatrix} U & 4t \\ 4t & -U \end{pmatrix} \right\}. \quad (75)$$

Rewriting the matrix

$$\begin{pmatrix} U & 4t \\ 4t & -U \end{pmatrix} = \sqrt{U^2 + 16t^2} \begin{pmatrix} \cos \Theta & \sin \Theta \\ \sin \Theta & -\cos \Theta \end{pmatrix}, \quad (76)$$

we find the ground state of the half-filled two-site Hubbard model

$$|\text{gs}\rangle = \cos \Theta/2 |\text{cov}_+\rangle + \sin \Theta/2 |\text{ion}_+\rangle \quad (77)$$

$$= \frac{1}{\sqrt{2}} \left(\cos \frac{\Theta}{2} c_{2\downarrow}^\dagger c_{1\uparrow}^\dagger + \cos \frac{\Theta}{2} c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger + \sin \frac{\Theta}{2} c_{1\downarrow}^\dagger c_{1\uparrow}^\dagger + \sin \frac{\Theta}{2} c_{2\downarrow}^\dagger c_{2\uparrow}^\dagger \right) | 0 \rangle \quad (78)$$

with an energy of $\varepsilon_{\text{gs}} = (U - \sqrt{U^2 + 16t^2})/2$. Without correlations ($U = 0 \leadsto \Theta = \pi/2$), all basis states have the same prefactor, so we can factorize the ground state, writing it as a product $c_{+\downarrow}^\dagger c_{+\uparrow}^\dagger | 0 \rangle$ of the operators defined in (71). For finite U this is no longer possible. In the strongly correlated limit $U \gg t$ ($\Theta \searrow 0$) the ground state becomes the maximally entangled state $|\text{cov}_+\rangle$

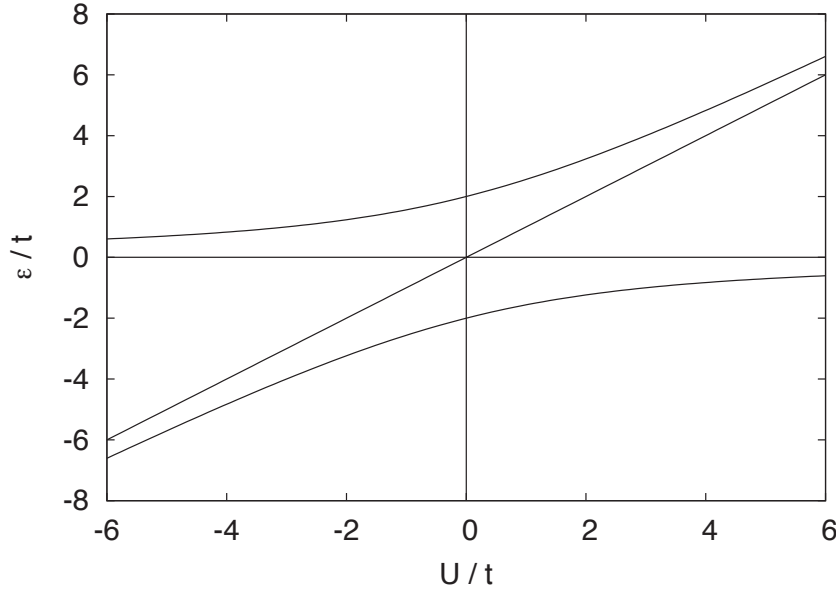


Fig. 5: Spectrum of the two-site Hubbard model as a function of U/t .

and can not even approximately be expressed as a two-electron Slater determinant. See [15] for a more detailed discussion, but beware that there the basis was chosen slightly differently to make the symmetry of the singlet/triplet state apparent.

We can, however, construct a product state, exploiting the freedom we gained by introducing second quantization: the product wave function in Fock space

$$|VB\rangle = \left(1 + c_{2\downarrow}^\dagger c_{1\uparrow}^\dagger\right) \left(1 + c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger\right) |0\rangle \quad (79)$$

$$= \underbrace{|0\rangle}_{N=0} + \underbrace{\left(c_{2\downarrow}^\dagger c_{1\uparrow}^\dagger + c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger\right) |0\rangle}_{N_\uparrow=1=N_\downarrow} + \underbrace{c_{2\downarrow}^\dagger c_{1\downarrow}^\dagger c_{2\uparrow}^\dagger c_{1\uparrow}^\dagger |0\rangle}_{N=4} \quad (80)$$

has a component in the two-electron Hilbert space that is just the covalent state $|\text{cov}_+\rangle$. It would be very desirable to generalize this approach to a half-filled state without double occupancies, i.e., a *Mott state* and to models with more than two sites. We might try an ansatz

$$|VB^?\rangle = \prod_{\langle ij \rangle} \left(1 + c_{j\downarrow}^\dagger c_{i\uparrow}^\dagger + c_{i\downarrow}^\dagger c_{j\uparrow}^\dagger\right) |0\rangle \quad (81)$$

that has the advantage of not producing doubly occupied sites. The product is over pairs of sites, i.e., bonds, where each site only occurs in one such bond (if a site i participated in two bonds $\langle ij \rangle$ and $\langle ik \rangle$, there would be terms with doubly occupied site i , e.g., $c_{k\downarrow}^\dagger c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger c_{k\uparrow}^\dagger$). There are, however, many ways we could partition the lattice sites into bonds, and to maintain the symmetry of the lattice we would have to sum over them. Alternatively, we could take products of all bond states and use a Gutzwiller projection to eliminate the doubly occupied sites. This is the idea of the *resonating valence bond* (RVB) state [20]. Unfortunately, neither approach to the Mott state is easy to handle.

In the *negative- U Hubbard model* we do not have such problems. For $U \ll t$ the ground state is a linear combination of doubly occupied sites $|\text{ion}_+\rangle$, which can be obtained from

$$|\text{pair}\rangle = \frac{1}{2} \left(1 + c_{2\downarrow}^\dagger c_{2\uparrow}^\dagger\right) \left(1 + c_{1\downarrow}^\dagger c_{1\uparrow}^\dagger\right) |0\rangle. \quad (82)$$

As each pair of creation operators in the product involves only a single site, this ansatz readily generalizes to larger lattices

$$|\text{pair}\rangle = \prod_i \frac{1}{\sqrt{2}} \left(1 + c_{i\downarrow}^\dagger c_{i\uparrow}^\dagger\right) |0\rangle. \quad (83)$$

The idea of such grand-canonical product states in Fock space are central for understanding the superconducting state.

5.2 BCS state

We now turn from the Hubbard model to the BCS Hamiltonian

$$H_{\text{BCS}} = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k,k'} V_{k,k'} c_{-k'\downarrow}^\dagger c_{k'\uparrow}^\dagger c_{k\uparrow} c_{-k\downarrow} \quad (84)$$

in which the interaction term scatters Cooper pairs of electrons ($k \uparrow, -k \downarrow$) with different values of k . We start again by looking at a two-site model. With periodic boundary conditions, the bonding and antibonding states (71) become states with $k = 0$ and $k = \pi$, respectively. Note that for both values ($k = \pi$ being at the boundary of the Brillouin zone) we have $k = -k$. Setting $V_{0,\pi} = -I$, we obtain the two-site Hamiltonian

$$H = \sum_{k \in \{0,\pi\}, \sigma} \varepsilon_k n_{k\sigma} - I \left(c_{\pi\downarrow}^\dagger c_{\pi\uparrow}^\dagger c_{0\uparrow} c_{0\downarrow} + c_{0\downarrow}^\dagger c_{0\uparrow}^\dagger c_{\pi\uparrow} c_{\pi\downarrow} \right). \quad (85)$$

For $N_\uparrow = 1 = N_\downarrow$ the Hamiltonian matrix is

$$\left\langle 0 \left| \begin{pmatrix} c_{0\uparrow} c_{\pi\downarrow} \\ c_{\pi\uparrow} c_{0\downarrow} \\ c_{0\uparrow} c_{0\downarrow} \\ c_{\pi\uparrow} c_{\pi\downarrow} \end{pmatrix} \right. H \left(\begin{pmatrix} c_{\pi\downarrow}^\dagger c_{0\uparrow}^\dagger & c_{0\downarrow}^\dagger c_{\pi\uparrow}^\dagger & c_{0\downarrow}^\dagger c_{0\uparrow}^\dagger & c_{\pi\downarrow}^\dagger c_{\pi\uparrow}^\dagger \end{pmatrix} \right| 0 \right\rangle = \begin{pmatrix} \varepsilon_0 + \varepsilon_\pi & 0 & 0 & 0 \\ 0 & \varepsilon_0 + \varepsilon_\pi & 0 & 0 \\ 0 & 0 & 2\varepsilon_0 & -I \\ 0 & 0 & -I & 2\varepsilon_\pi \end{pmatrix} \quad (86)$$

To find the ground state, we need only consider the subspace of the Cooper pairs

$$|\text{pair}_k\rangle = c_{-k\downarrow}^\dagger c_{k\uparrow}^\dagger |0\rangle. \quad (87)$$

Writing $\bar{\varepsilon} = (\varepsilon_0 + \varepsilon_\pi)/2$ and $\Delta = \varepsilon_\pi - \varepsilon_0$

$$\begin{pmatrix} 2\varepsilon_0 & -I \\ -I & 2\varepsilon_\pi \end{pmatrix} = 2\bar{\varepsilon} - \begin{pmatrix} \Delta & I \\ I & -\Delta \end{pmatrix} = 2\bar{\varepsilon} - \sqrt{I^2 + \Delta^2} \begin{pmatrix} \cos \Theta & \sin \Theta \\ \sin \Theta & -\cos \Theta \end{pmatrix} \quad (88)$$

we diagonalize, just as we did in the case of the Hubbard model, to find the ground state of the half-filled two-site BCS-model for $I > 0$

$$|\text{gs}\rangle = \cos \Theta/2 |\text{pair}_0\rangle + \sin \Theta/2 |\text{pair}_\pi\rangle = \left(\cos \frac{\Theta}{2} c_{0\downarrow}^\dagger c_{0\uparrow}^\dagger + \sin \frac{\Theta}{2} c_{\pi\downarrow}^\dagger c_{\pi\uparrow}^\dagger \right) |0\rangle \quad (89)$$

It is similar in form to the ground state (82) of the negative- U two-site Hubbard model, except that the two pairs can have different amplitudes, as the pair with lower band energy ε_k is preferred. Introducing $\Theta_0 = \Theta$ and $\Theta_\pi = \pi - \Theta$ we can recover this state (for any $I > 0$, not just in the limit of large interaction as for the negative- U Hubbard model) from the Fock-space product-state

$$|\text{BCS}\rangle = \prod_{k \in \{0, \pi\}} \frac{1}{\sqrt{1 + \cos^2 \frac{\Theta_k}{2}}} \left(1 + \cos \frac{\Theta_k}{2} c_{-k\downarrow}^\dagger c_{k\uparrow}^\dagger \right) |0\rangle. \quad (90)$$

This readily generalizes to larger numbers of k -points, where it becomes the BCS wave function.

6 Conclusions

We have studied the consequences of one of the most bizarre features of quantum mechanics, the existence of indistinguishable particles. To treat such particles, we have to introduce artificial labels but must make sure that no observable depends on them. The invariance under permutations of these labels implies that many-particle wave functions must be properly (anti)symmetrized. The type of symmetry is given by the spin-statistics connection. Unfortunately, imposing the correct (anti)symmetry on a generic N -particle wave function is a computationally hard problem as there are $N!$ permutations. One way to get around this problem is to integrate-out all degrees of freedom that are not explicitly considered. This gives rise to the reduced density matrices. Another is to exploit the fact that products of single-particle wave functions can be efficiently anti-symmetrized by forming the Slater determinant. Working with Slater determinants is made more convenient by introducing operators that are designed to encode the Fermi sign in their position. This technique of second quantization has two important benefits: we are no longer restricted to calculating with Slater determinants in configuration-space representation but can work with abstract Dirac states instead. Even more importantly, creation and annihilation operators are defined in Fock space. They enable us to write observables in a unified way on Fock space. Moreover, they allow us to also write wave functions in Fock space. Using this additional degree of freedom, it is possible to write non-Fermi-liquid states as generalized Slater determinants (product states), the most famous being the BCS state.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] P. Forrest: *The Identity of Indiscernibles*, in The Stanford Encyclopedia of Philosophy (Winter 2012 Ed.), E.N. Zalta (ed.)
<http://plato.stanford.edu/entries/identity-indiscernible/>
- [2] O. Carnal and J. Mlynek: *Young's Double-Slit Experiment with Atoms*, Phys. Rev. Lett. **66**, 2689 (1991)
- [3] W. Pauli: *The connection between spin and statistics*, Phys. Rev. **58**, 715 (1940)
- [4] R.F. Streater and A.S. Wightman: *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964)
- [5] R.P. Feynman: *The reason for antiparticles*, in R. MacKenzie and P. Doust (eds.): *Elementary Particles and the Laws of Physics* (Cambridge University Press, 1987)
- [6] S. Tomonaga: *The Story of Spin* (University of Chicago Press, 1997)
- [7] M. Berry and J. Robbins: *Quantum Indistinguishability: Spin-statistics without Relativity or Field Theory?*, in R.C. Hilborn and G.M. Tino (eds.): *Spin-Statistics Connection and Commutation Relations*, AIP Conference Proceedings Vol. 545, pp. 3-13 (2000)
- [8] J.M. Leinaas and J. Myrheim: *On the theory of identical particles*, Il Nuovo Cimento B **37** 1 (1977)
- [9] D. Ceperley: *Fermion Nodes*, J. Stat. Phys. **63**, 1237 (1991)
- [10] E.H. Lieb and F.Y. Wu: *The one-dimensional Hubbard model*, Physica A **321**, 1 (2003)
- [11] P.-O. Löwdin: *Quantum Theory of Many-Particle Systems I*, Phys. Rev. **97**, 1474 (1955)
- [12] A.J. Coleman and V.I. Yukalov: *Reduced Density Matrices: Coulson's Challenge* (Springer, New York, 2001)
- [13] P. Gori-Giorgi, F. Sacchetti, and G.B. Bachelet: *Analytic static structure factors and pair-correlation functions for the unpolarized homogeneous electron gas*, Phys. Rev. B **61**, 7353 (2000)
- [14] A.W. Overhauser: *Structure of nuclear matter*, Phys. Rev. Lett. **4**, 415 (1960)
- [15] E. Koch: *Exchange Mechanisms*, in [21]
- [16] P.M.A. Dirac: *The Quantum Theory of Emission and Absorption of Radiation*, Proc. Roy. Soc. A **114**, 243 (1927)
- [17] P. Jordan and O. Klein: *Zum Mehrkörperproblem in der Quantenmechanik*, Z. Physik **45**, 751 (1927)

- [18] P. Jordan and E. Wigner: *Über das Paulische Äquivalenzverbot*, Z. Physik **47**, 631 (1928)
- [19] R. Eder: *Multiplets in Transition Metal Ions*, in [21]
- [20] P.W. Anderson; *Resonating valence bonds: A new kind of insulator?*
Materials Research Bulletin **8**, 153 (1973)
- [21] E. Pavarini, E. Koch, F. Anders, and M. Jarrell (eds.):
Correlated Electrons: From Models to Materials
Reihe Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
<http://www.cond-mat.de/events/correl12>

3 Magnetism: Models and Mechanisms

Eva Pavarini

Institute for Advanced Simulation

Forschungszentrum Jülich

Contents

1	Magnetism in strongly-correlated systems	2
2	The Hubbard model	7
2.1	Itinerant magnetism	10
2.2	Isolated magnetic ions	17
2.3	Interacting localized moments	21
3	The Kondo model	28
4	Conclusion	31
A	Formalism	33
A.1	Matsubara Green functions	33
A.2	Linear response theory	36
A.3	Magnetic susceptibility	39

1 Magnetism in strongly-correlated systems

Long-range magnetic order is a manifestation of emergence, the hallmark of strong electron-electron correlations. It arises from the same interactions that lead to the metal-insulator transition and orbital-ordering or that give rise to the Kondo effect. And yet, magnetic order phenomena can, to a large extent, be explained by solving spin models and forgetting about the microscopic mechanisms which justify them. To understand models and mechanisms we have, however, to take a step back into the complex world of strong correlations [1–7].

Magnetism ultimately arises from the intrinsic magnetic moment of electrons, $\boldsymbol{\mu} = -g\mu_B\mathbf{s}$, where μ_B is the Bohr magneton and $g \simeq 2.0023$ is the electronic g -factor. It is however an inherently quantum mechanical effect, the consequence of the interplay between Pauli exclusion principle, Coulomb electron-electron interaction, and hopping of electrons. To understand this let us consider the simplest possible system, an isolated atom or ion. In the non-relativistic limit electrons in a single ion are typically described by the Hamiltonian

$$H_e^{\text{NR}} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \frac{Z}{r_i} + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|},$$

where Z is the atomic number and $\{\mathbf{r}_i\}$ are the coordinates of the electrons with respect to the ionic nucleus. Here, as in the rest of this lecture, we use atomic units. If we consider only the external atomic shell with quantum numbers nl , for example the $3d$ shell of transition-metal ions, we can rewrite this Hamiltonian as follows

$$H_e^{\text{NR}} = \varepsilon_{nl} \sum_{m\sigma} c_{m\sigma}^\dagger c_{m\sigma} + \frac{1}{2} \sum_{\sigma\sigma'} \sum_{m\tilde{m}m'\tilde{m}'} U_{mm'\tilde{m}\tilde{m}'}^l c_{m\sigma}^\dagger c_{m'\sigma'}^\dagger c_{\tilde{m}'\sigma'} c_{\tilde{m}\sigma}. \quad (1)$$

Here ε_{nl} is the energy of the electrons in the nl atomic shell and m the degenerate one-electron states in that shell. For a hydrogen-like atom

$$\varepsilon_{nl} = -\frac{1}{2} \frac{Z^2}{n^2}.$$

The couplings $U_{mm'\tilde{m}\tilde{m}'}^l$ are the four-index Coulomb integrals. In a basis of atomic functions the bare Coulomb integrals are

$$U_{mm'\tilde{m}\tilde{m}'}^{ijj'j'} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}}(\mathbf{r}_1) \overline{\psi_{jm'\sigma'}}(\mathbf{r}_2) \psi_{j'\tilde{m}'\sigma'}(\mathbf{r}_2) \psi_{i\tilde{m}\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

and

$$U_{mm'\tilde{m}\tilde{m}'}^l = U_{mm'\tilde{m}\tilde{m}'}^{iiii} \quad m, m', \tilde{m}, \tilde{m}' \in nl \text{ shell}.$$

The eigenstates of Hamiltonian (1) for fixed number of electrons, N , are the multiplets [8, 9]. Since in H_e^{NR} the Coulomb repulsion and the central potential are the only interactions, the multiplets can be labeled with S and L , the quantum numbers of the electronic total spin and total orbital angular momentum operators, $\mathbf{S} = \sum_i \mathbf{s}_i$, and $\mathbf{L} = \sum_i \mathbf{l}_i$. Closed-shell ions have $S = L = 0$ in their ground state. Ions with a partially-filled shell are called *magnetic ions*; the value of S and L for their ground state can be obtained via two rules due to Friedrich Hund. They say that the lowest-energy multiplet is the one with

- the largest value of S
- the largest value of L compatible with the previous rule

The main relativistic effect is the spin-orbit interaction, which has the form $H_e^{\text{SO}} = \sum_i \lambda_i \mathbf{l}_i \cdot \mathbf{s}_i$. For not too heavy atoms it is a weak perturbation. Then, for electrons in a given shell, we can use the first and second Hund's rule to rewrite H_e^{SO} in a simpler form

$$H_e^{\text{SO}} \sim \lambda \mathbf{L} \cdot \mathbf{S} = \frac{1}{2} \lambda (\mathbf{J}^2 - \mathbf{S}^2 - \mathbf{L}^2), \quad (2)$$

$$\lambda \sim [2\Theta(1 - 2n) - 1] g\mu_B^2 \frac{1}{2S} \left\langle \frac{1}{r} \frac{d}{dr} v_R(r) \right\rangle,$$

where n is the filling and Θ the step function; $v_R(r)$ is the effective potential, which includes, e.g., the Hartree electron-electron term [10]. For a hydrogen-like atom, $v_R(r) = -Z/r$. Because of the LS coupling (2) the eigenstates have quantum numbers L , S and J , where $\mathbf{J} = \mathbf{S} + \mathbf{L}$ is the total angular momentum. The value of J in the ground-state multiplet is given by the third Hund's rule

$$\bullet \text{ total angular momentum } J = \begin{cases} |L - S| & \text{for filling } n < 1/2 \\ S & \text{for filling } n = 1/2 \\ L + S & \text{for filling } n > 1/2 \end{cases}$$

In the presence of spin-orbit interaction a given multiplet is then labeled by $^{2S+1}L_J$, and its states can be indicated as $|JJ_zLS\rangle$. If we consider, e.g., the case of the ion Cu^{2+} , characterized by the $[\text{Ar}] 3d^9$ electronic configuration, Hund's rules tell us that the $3d$ ground-state multiplet has quantum numbers $S = 1/2$, $L = 2$ and $J = 5/2$. A Mn^{3+} ion, which is in the $[\text{Ar}] 3d^4$ electronic configuration, has instead a ground-state multiplet with quantum numbers $S = 2$, $L = 2$ and $J = 0$. The order of the Hund's rules reflects the hierarchy of the interactions. The strongest interactions are the potential $v_R(r)$, which determines ε_{nl} , and the average Coulomb interaction, the strength of which is measured by the average *direct Coulomb integral*,

$$U_{\text{avg}} = \frac{1}{(2l+1)^2} \sum_{mm'} U_{mm'mm'}^l.$$

For a N -electron state the energy associated with these two interactions is $E(N) = \varepsilon_{nl}N + U_{\text{avg}}N(N-1)/2$, the same for all multiplets of a given shell. The first Hund's rule is instead due to the average *exchange Coulomb integral*, J_{avg} , defined as

$$U_{\text{avg}} - J_{\text{avg}} = \frac{1}{2l(2l+1)} \sum_{mm'} (U_{mm'mm'}^l - U_{mm'm'm}^l),$$

which is the second largest Coulomb term; for transition-metal ions $J_{\text{avg}} \sim 1$ eV. Smaller Coulomb integrals determine the orbital anisotropy of the Coulomb matrix and the second Hund's rule.¹ The third Hund's rule comes, as we have seen, from the spin-orbit interaction which, for not too heavy atoms, is significantly weaker than all the rest.

¹For more details on Coulomb integrals and their averages see Ref. [10].

The role of Coulomb electron-electron interaction in determining S and L can be understood through the simple example of a C atom, electronic configuration [He] $2s^2 2p^2$. We consider only the p shell, filled by two electrons. The Coulomb exchange integrals have the form

$$\begin{aligned} J_{m,m'}^p &= U_{mm'm'm}^p \\ &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}(\mathbf{r}_1)} \overline{\psi_{im'\sigma}(\mathbf{r}_2)} \psi_{im\sigma}(\mathbf{r}_2) \psi_{im'\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} \\ &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\phi_{imm'\sigma}(\mathbf{r}_1) \overline{\phi_{imm'\sigma}(\mathbf{r}_2)}}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{V} \sum_{\mathbf{k}} \frac{4\pi}{k^2} |\phi_{imm'\sigma}(\mathbf{k})|^2, \end{aligned} \quad (3)$$

and they are therefore positive. They generate the Coulomb-interaction term

$$-\frac{1}{2} \sum_{\sigma} \sum_{m \neq m'} J_{m,m'}^p c_{m\sigma}^{\dagger} c_{m\sigma} c_{m'\sigma}^{\dagger} c_{m'\sigma} = -\frac{1}{2} \sum_{m \neq m'} 2J_{m,m'}^p \left[S_z^m S_z^{m'} + \frac{1}{4} n_m n_{m'} \right].$$

This interaction yields an *energy gain* if the two electrons occupy two different p orbitals with parallel spins, hence favors the state with the largest spin (first Hund's rule). It turns out that for the p^2 configuration there is only one possible multiplet with $S = 1$, and such a state has $L = 1$. There are instead two excited $S = 0$ multiplets, one with $L = 0$ and the other with $L = 2$; the latter is the one with the lowest energy (second Hund's rule).

To understand the magnetic properties of an isolated ion we have to analyze how its levels are modified by an external magnetic field \mathbf{h} . The effect of a magnetic field is described by

$$H_e^H = \mu_B (g\mathbf{S} + \mathbf{L}) \cdot \mathbf{h} + \frac{\hbar^2}{8} \sum_i (x_i^2 + y_i^2) = H_e^Z + H_e^L. \quad (4)$$

The linear term is the Zeeman Hamiltonian. If the ground-state multiplet is characterized by $J \neq 0$ the Zeeman interaction splits its $2J + 1$ degenerate levels. The second order term yields Larmor diamagnetism, which is usually only important if the ground-state multiplet has $J = 0$, as it happens for ions with closed external shells. The energy $\mu_B h$ is typically very small (for a field as large as 100 T it is as small as 6 meV); it can however be comparable with or larger than the spin-orbit interaction if the latter is tiny (very light atoms). Taking all interactions into account, the total Hamiltonian is

$$H_e \sim H_e^{\text{NR}} + H_e^{\text{SO}} + H_e^H.$$

In a crystal the electronic Hamiltonian is complicated by the interaction with other nuclei and their electrons. The non-relativistic part of the Hamiltonian takes then the form

$$H_e^{\text{NR}} = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq i'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} - \sum_{i\alpha} \frac{Z_{\alpha}}{|\mathbf{r}_i - \mathbf{R}_{\alpha}|} + \frac{1}{2} \sum_{\alpha \neq \alpha'} \frac{Z_{\alpha} Z_{\alpha'}}{|\mathbf{R}_{\alpha} - \mathbf{R}_{\alpha'}|},$$

where Z_{α} is the atomic number of the nucleus located at position \mathbf{R}_{α} . In a basis of localized Wannier functions [10] this Hamiltonian can be written as

$$\begin{aligned} H_e^{\text{NR}} &= - \sum_{ii'\sigma} \sum_{mm'} t_{m,m'}^{i,i'} c_{im\sigma}^{\dagger} c_{i'm'\sigma} \\ &\quad + \frac{1}{2} \sum_{ii'jj'} \sum_{\sigma\sigma'} \sum_{mm'} \sum_{\tilde{m}\tilde{m}'} U_{mm'\tilde{m}\tilde{m}'}^{ijj'} c_{im\sigma}^{\dagger} c_{jm'\sigma'}^{\dagger} c_{j'\tilde{m}'\sigma'} c_{i'\tilde{m}\sigma}, \end{aligned} \quad (5)$$

where

$$t_{m,m'}^{i,i'} = - \int d\mathbf{r} \overline{\psi_{im\sigma}(\mathbf{r})} \left[-\frac{1}{2} \nabla^2 + v_R(\mathbf{r}) \right] \psi_{i'm'\sigma}(\mathbf{r}).$$

The terms $\varepsilon_{m,m'} = -t_{m,m'}^{i,i}$ yield the crystal-field matrix and $t_{m,m'}^{i,i'}$ with $i \neq i'$ the hopping integrals. The label m indicates here the orbital quantum number of the Wannier function. In general the Hamiltonian (5) will include states stemming from more than a single atomic shell. For example, in the case of strongly-correlated transition-metal oxides, the set $\{im\}$ includes transition-metal $3d$ and oxygen $2p$ states. The exact solution of the many-body problem described by (5) is an impossible challenge. The reason is that the properties of a many-body system are inherently emergent and hence hard to predict *ab-initio* in the lack of any understanding of the mechanism behind them. In this lecture, however, we want to focus on magnetism. Since the nature of cooperative magnetic phenomena in crystals is nowadays to a large extent understood, we can find realistic approximations to (5) and even map it onto simpler models which still retain the essential ingredients to explain long-range magnetic order.

Let us identify the parameters of the electronic Hamiltonian important for magnetism. The first is the crystal-field matrix $\varepsilon_{m,m'}$. The crystal field at a given site i is a non-spherical potential due to the joint effect of the electric field generated by the surrounding ions and of covalent-bond formation [9]. The crystal field might split the levels within a given shell and has therefore a strong impact on magnetic properties. We can identify three ideal regimes. In the *strong crystal field* limit the crystal field splitting is so large that it is comparable with the average Coulomb exchange responsible for the first Hund's rule. This can happen in $4d$ or $5d$ transition-metal oxides. A consequence of an *intermediate crystal field* (weaker than the average Coulomb exchange but larger than Coulomb anisotropy and spin-orbit interaction) is the quenching of the angular momentum, $\langle \mathbf{L} \rangle = 0$. In this limit the second and third Hund's rule are not respected. This typically happens in $3d$ transition-metal oxides. In $4f$ systems the crystal-field splitting is usually much weaker than the spin-orbit coupling (*weak crystal field* limit) and mainly splits states within a given multiplet, leaving a reduced magnetic moment. In all three cases, because of the crystal field, a magnetic ion in a crystal might lose, totally or partially, its spin, angular or total moment. Or, sometimes, it is the other way around. This happens for Mn^{3+} ions, which should have a $J = 0$ ground state according to the third Hund's rule. However in perovskites such as LaMnO_3 they behave as $S = 2$ ions because of the quenching of the angular momentum. Even if the crystal field does not suppress the magnetic moment of the ion, the electrons might delocalize to form broad bands completely losing their original atomic character. This happens, e.g., if the hopping integrals $t_{m,m'}^{i,i'}$ are much larger than the average on-site Coulomb interaction U_{avg} . Surprisingly, magnetic instabilities arise even in the absence of localized moments. This *itinerant magnetism* is mostly due to band effects, i.e., it is associated with a large one-electron linear static response-function, $\chi_0(\mathbf{q}; 0)$. In this limit correlation effects are typically weak. To study them we can exploit the power of the *standard model* of solid-state physics, the density-functional theory (DFT), taking into account Coulomb interaction effects beyond the local-density approximation (LDA) at the perturbative level, e.g., in the random-phase approximation (RPA). With this approach we can understand and describe Stoner instabilities.

In the opposite limit, the *local moments* regime, the hopping integrals are small with respect to U_{avg} . This is the regime of strong electron-electron correlations, where complex many-body effects, e.g., those leading to the Mott metal-insulator transition, play an important role. At *low enough energy*, however, only spin excitations matter. Ultimately, at integer filling we can integrate out (*downfold*) charge fluctuations and describe the system via effective spin Hamiltonians. The latter typically take the form

$$H_S = \frac{1}{2} \sum_{ii'} \Gamma^{i,i'} \mathbf{S}_i \cdot \mathbf{S}_{i'} + \dots = H_S^H + \dots \quad (6)$$

The term H_S^H given explicitly in (6) is the Heisenberg Hamiltonian, and $\Gamma^{i,i'}$ is the Heisenberg exchange coupling, which can be antiferromagnetic ($\Gamma^{i,i'} > 0$) or ferromagnetic ($\Gamma^{i,i'} < 0$). The Hamiltonian (6) can, for a specific system, be quite complicated, and might include long-range exchange interactions or anisotropic terms. Nevertheless, it represents a huge simplification compared to the unsolvable many-body problem described by (5), since, at least within very good approximated schemes, it can be solved. Spin Hamiltonians of type (6) are the minimal models which still provide a realistic picture of long-range magnetic order in strongly-correlated insulators. There are various sources of exchange couplings. Electron-electron repulsion itself yields via Coulomb exchange a ferromagnetic Heisenberg interaction, the *Coulomb exchange interaction*. The origin of such interaction can be understood via a simple model with a single orbital, m . The inter-site Coulomb exchange coupling has then the form

$$J^{i,i'} = U_{mmmm}^{ii' i' i} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\overline{\psi_{im\sigma}(\mathbf{r}_1)} \overline{\psi_{i'm\sigma}(\mathbf{r}_2)} \psi_{im\sigma}(\mathbf{r}_2) \psi_{i'm\sigma}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

and it is therefore positive, as one can show by following the same steps that we used in Eq. (3) for $J_{m,m'}^p$. Hence, the corresponding Coulomb interaction yields a ferromagnetic Heisenberg-like Hamiltonian with $\Gamma^{i,i'} = -2J^{i,i'} < 0$. A different source of magnetic interactions are the *kinetic exchange* mechanisms (direct exchange, super-exchange, double exchange, Rudermann-Kittel-Kasuya-Yosida interaction ...), which are mediated by the hopping integrals. Kinetic exchange couplings are typically (with few well understood exceptions) antiferromagnetic [11]. A representative example of kinetic exchange will be discussed in the next section.

While itinerant and local moment regime are very interesting ideal limit cases, correlated materials elude rigid classifications. The same system can present features associated with both regimes, although at different temperatures and/or energy scales. This happens in Kondo systems, heavy Fermions, metallic strongly-correlated materials, and doped Mott insulators.

In this lecture we will discuss in representative cases the itinerant and localized moment regime and their crossover, as well as the most common mechanisms leading to magnetic cooperative phenomena. Since our target is to understand strongly-correlated materials, we adopt the formalism typically used for these systems. A concise introduction to Matsubara Green functions, correlation functions, susceptibilities and linear-response theory can be found in the Appendix.

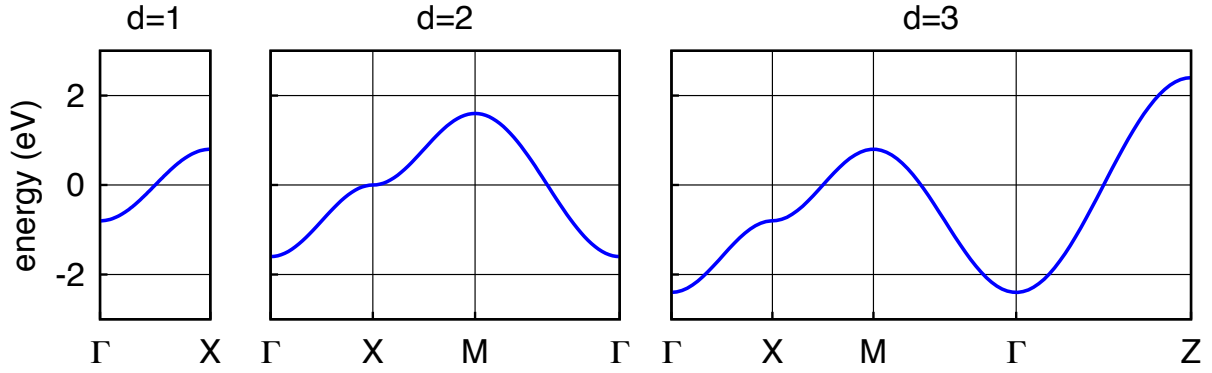


Fig. 1: The band structure of the one-band tight-binding model (hypercubic lattice). The hopping integral is $t = 0.4$ eV. From left to right: one-, two-, and three-dimensional case. At half-filling ($n = 1$) the Fermi level is at zero energy.

2 The Hubbard model

The simplest model that we can consider is the one-band Hubbard model

$$H = \varepsilon_d \sum_i \sum_{\sigma} c_{i\sigma}^{\dagger} c_{i\sigma} - t \sum_{\langle ii' \rangle} \sum_{\sigma} c_{i\sigma}^{\dagger} c_{i'\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} = H_d + H_T + H_U, \quad (7)$$

where ε_d is the on-site energy, t is the hopping integral between first nearest neighbors $\langle ii' \rangle$ and U the on-site Coulomb repulsion; $c_{i\sigma}^{\dagger}$ creates an electron in a Wannier state with spin σ centered at site i and $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$. The Hubbard model is a simplified version of Hamiltonian (5) with $m = m' = \tilde{m} = \tilde{m}' = 1$ and

$$\begin{cases} \varepsilon_d &= -t_{1,1}^{i,i} \\ t &= t_{1,1}^{i,i'} \\ U &= U_{1111}^{iiii} \end{cases}.$$

In the $U = 0$ limit the Hubbard model describes a system of independent electrons. The Hamiltonian is then diagonal in the Bloch basis

$$H_d + H_T = \sum_{\mathbf{k}} \sum_{\sigma} [\varepsilon_d + \varepsilon_{\mathbf{k}}] c_{\mathbf{k}\sigma}^{\dagger} c_{\mathbf{k}\sigma}. \quad (8)$$

The energy dispersion $\varepsilon_{\mathbf{k}}$ depends on the geometry and dimensionality d of the lattice. For a hypercubic lattice in d dimensions

$$\varepsilon_{\mathbf{k}} = -2t \sum_{\nu=1}^d \cos(k_{r_{\nu}} a),$$

where a is the lattice constant, and $r_1 = x, r_2 = y, r_3 = z$. The energy $\varepsilon_{\mathbf{k}}$ does not depend on the spin. In Fig. 1 we show $\varepsilon_{\mathbf{k}}$ in the one-, two- and three-dimensional case.

In the opposite limit ($t = 0$) the Hubbard model describes a collection of isolated atoms. Each atom has four electronic many-body states

$ N, S, S_z\rangle$		N	S	$E(N)$	
$ 0, 0, 0\rangle$	$= 0\rangle$	0	0	0	
$ 1, \frac{1}{2}, \uparrow\rangle$	$= c_{i\uparrow}^\dagger 0\rangle$	1	1/2	ε_d	(9)
$ 1, \frac{1}{2}, \downarrow\rangle$	$= c_{i\downarrow}^\dagger 0\rangle$	1	1/2	ε_d	
$ 2, 0, 0\rangle$	$= c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger 0\rangle$	2	0	$2\varepsilon_d + U$	

where $E(N)$ is the total energy, N the total number of electrons and S the total spin. We can express the atomic Hamiltonian $H_d + H_U$ in a form in which the dependence on N_i , S_i , and S_z^i is explicitly given

$$H_d + H_U = \varepsilon_d \sum_i n_i + U \sum_i \left[- (S_z^i)^2 + \frac{n_i^2}{4} \right], \quad (10)$$

where $S_z^i = (n_{i\uparrow} - n_{i\downarrow})/2$ is the z component of the spin operator and $n_i = \sum_\sigma n_{i\sigma} = N_i$.

In the large t/U limit and at half-filling we can downfold charge fluctuations and map the Hubbard model into an effective spin model of the form

$$H_S = \frac{1}{2} \Gamma \sum_{\langle ii' \rangle} \left[\mathbf{S}_i \cdot \mathbf{S}_{i'} - \frac{1}{4} n_i n_{i'} \right]. \quad (11)$$

The coupling Γ can be calculated by using second-order perturbation theory. For a state in which two neighbors have opposite spins, $|\uparrow, \downarrow\rangle = c_{i\uparrow}^\dagger c_{i'\downarrow}^\dagger |0\rangle$, we obtain the energy gain

$$\Delta E_{\uparrow\downarrow} \sim - \sum_I \langle \uparrow, \downarrow | H_T | I \rangle \langle I | \frac{1}{E(2) + E(0) - 2E(1)} | I \rangle \langle I | H_T | \uparrow, \downarrow \rangle \sim - \frac{2t^2}{U}.$$

Here $|I\rangle$ ranges over the excited states with one of the two neighboring sites doubly occupied and the other empty, $|\uparrow\downarrow, 0\rangle = c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger |0\rangle$, or $|0, \uparrow\downarrow\rangle = c_{i'\uparrow}^\dagger c_{i'\downarrow}^\dagger |0\rangle$; these states can be occupied via virtual hopping processes. For a state in which two neighbors have parallel spins, $|\uparrow, \uparrow\rangle = c_{i\uparrow}^\dagger c_{i'\uparrow}^\dagger |0\rangle$, no virtual hopping is possible because of the Pauli principle, and $\Delta E_{\uparrow\uparrow} = 0$. Thus

$$\frac{1}{2} \Gamma \sim (\Delta E_{\uparrow\uparrow} - \Delta E_{\uparrow\downarrow}) = \frac{1}{2} \frac{4t^2}{U}. \quad (12)$$

The exchange coupling $\Gamma = 4t^2/U$ is positive, i.e., antiferromagnetic.

Canonical transformations [12] provide a scheme to derive systematically the effective spin model at any perturbation order. Let us consider a unitary transformation of the Hamiltonian

$$H_S = e^{iS} H e^{-iS} = H + [iS, H] + \frac{1}{2} [iS, [iS, H]] + \dots$$

We search for a transformation operator which eliminates, at a given order, hopping integrals between states with a different number of doubly occupied states. To do this first we split the

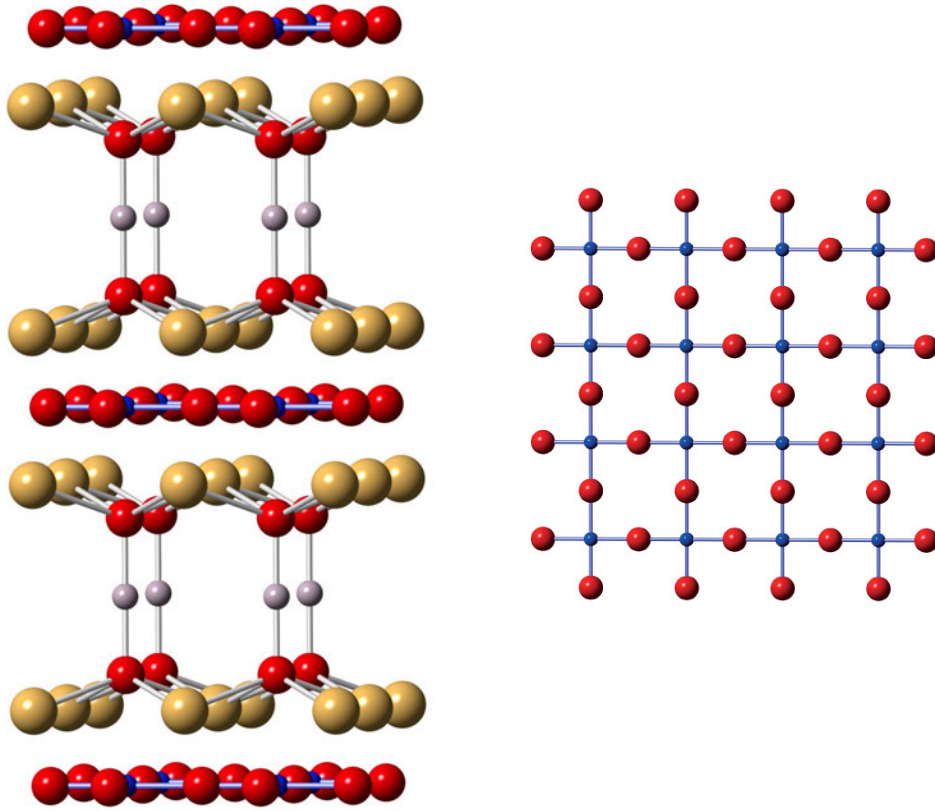


Fig. 2: Left: The crystal structure of $\text{HgBa}_2\text{CuO}_4$ showing the two-dimensional CuO_2 layers. Spheres represent atoms of Cu (blue), O (red), Ba (yellow), and Hg (grey). Right: A CuO_2 layer. The first nearest-neighbors hopping integral between neighboring Cu sites, t , is roughly given by $\sim 4t_{pd}^2/\Delta_{dp}$, where t_{pd} is the hopping between Cu d and O p states and $\Delta_{dp} = \varepsilon_d - \varepsilon_p$ their charge-transfer energy.

kinetic term H_T into a component, H_T^0 , which does not change the number of doubly occupied states and two terms which either increase it (H_T^+) or decrease it (H_T^-) by one

$$H_T = -t \sum_{\langle ii' \rangle} \sum_{\sigma} c_{i\sigma}^\dagger c_{i'\sigma} = H_T^0 + H_T^+ + H_T^-,$$

where

$$\begin{aligned} H_T^0 &= -t \sum_{\langle ii' \rangle} \sum_{\sigma} n_{i-\sigma} c_{i\sigma}^\dagger c_{i'\sigma} n_{i'-\sigma} \\ &\quad -t \sum_{\langle ii' \rangle} \sum_{\sigma} (1 - n_{i-\sigma}) c_{i\sigma}^\dagger c_{i'\sigma} (1 - n_{i'-\sigma}), \end{aligned}$$

$$H_T^+ = -t \sum_{\langle ii' \rangle} \sum_{\sigma} n_{i-\sigma} c_{i\sigma}^\dagger c_{i'\sigma} (1 - n_{i'-\sigma}),$$

$$H_T^- = (H_T^+)^\dagger.$$

The term H_T^0 commutes with H_U . The remaining two terms fulfill the commutation rules

$$[H_T^\pm, H_U] = \mp U H_T^\pm.$$

The operator S can be expressed as a linear combination of powers of the three operators H_T^0 , H_T^+ , and H_T^- . The actual combination which gives the effective spin model at a given order can be found via a recursive procedure [12]. At half-filling and second order, however, we can simply guess the form of S which leads to the Hamiltonian (11). By defining

$$S = -\frac{i}{U} (H_T^+ - H_T^-)$$

we obtain

$$H_S = H_U + H_T^0 + \frac{1}{U} \{ [H_T^+, H_T^-] + [H_T^0, H_T^-] + [H_T^+, H_T^0] \} + O(U^{-2}).$$

If we restrict the Hilbert space of H_S to the subspace with one electron per site (half filling), no hopping is possible without increasing the number of occupied states; hence, only the term $H_T^- H_T^+$ contributes. After some algebra, we obtain $H_S = H_S^{(2)} + O(U^{-2})$ with

$$H_S^{(2)} = \frac{1}{2} \frac{4t^2}{U} \sum_{ii'} \left[\mathbf{S}_i \cdot \mathbf{S}_{i'} - \frac{1}{4} n_i n_{i'} \right].$$

The Hubbard model (7) is seldom realized in Nature in this form. To understand real materials one typically has to take into account orbital degrees of freedom, long-range hopping integrals and sometimes longer range Coulomb interactions or perhaps even more complex many-body terms. Nevertheless, there are very interesting systems whose low-energy properties are, in the first approximation, described by (7). These are strongly correlated organic crystals (one-dimensional case) and high-temperature superconducting cuprates, in short HTSCs (two-dimensional case). An example of HTSC is $\text{HgBa}_2\text{CuO}_4$, whose structure is shown in Fig. 2. It is made of CuO_2 planes well divided by BaO-Hg-BaO blocks. The $x^2 - y^2$ -like states stemming from the CuO_2 planes can be described via a one-band Hubbard model. The presence of a $x^2 - y^2$ -like band at the Fermi level is a common feature of all HTSCs.

2.1 Itinerant magnetism

2.1.1 Pauli paramagnetism

Let us consider first the non-interacting limit of the Hubbard model, Hamiltonian (8). In the presence of an external magnetic field $\mathbf{h} = h_z \hat{z}$ the energy $\varepsilon_{\mathbf{k}}$ of a Bloch state is modified by the Zeeman interaction (4) as follows

$$\varepsilon_{\mathbf{k}} \rightarrow \varepsilon_{\mathbf{k}\sigma} = \varepsilon_{\mathbf{k}} + \frac{1}{2} \sigma g \mu_B h_z,$$

where we take the direction of the magnetic field as quantization axis and where on the right-hand side $\sigma = 1$ or -1 depending if the spin is parallel or antiparallel to \mathbf{h} . Thus, at linear order in the magnetic field, the $T = 0$ magnetization of the system is

$$M_z = -\frac{1}{2} (g \mu_B) \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} [n_{\mathbf{k}\uparrow} - n_{\mathbf{k}\downarrow}] \sim \frac{1}{4} (g \mu_B)^2 \rho(\varepsilon_F) h_z,$$

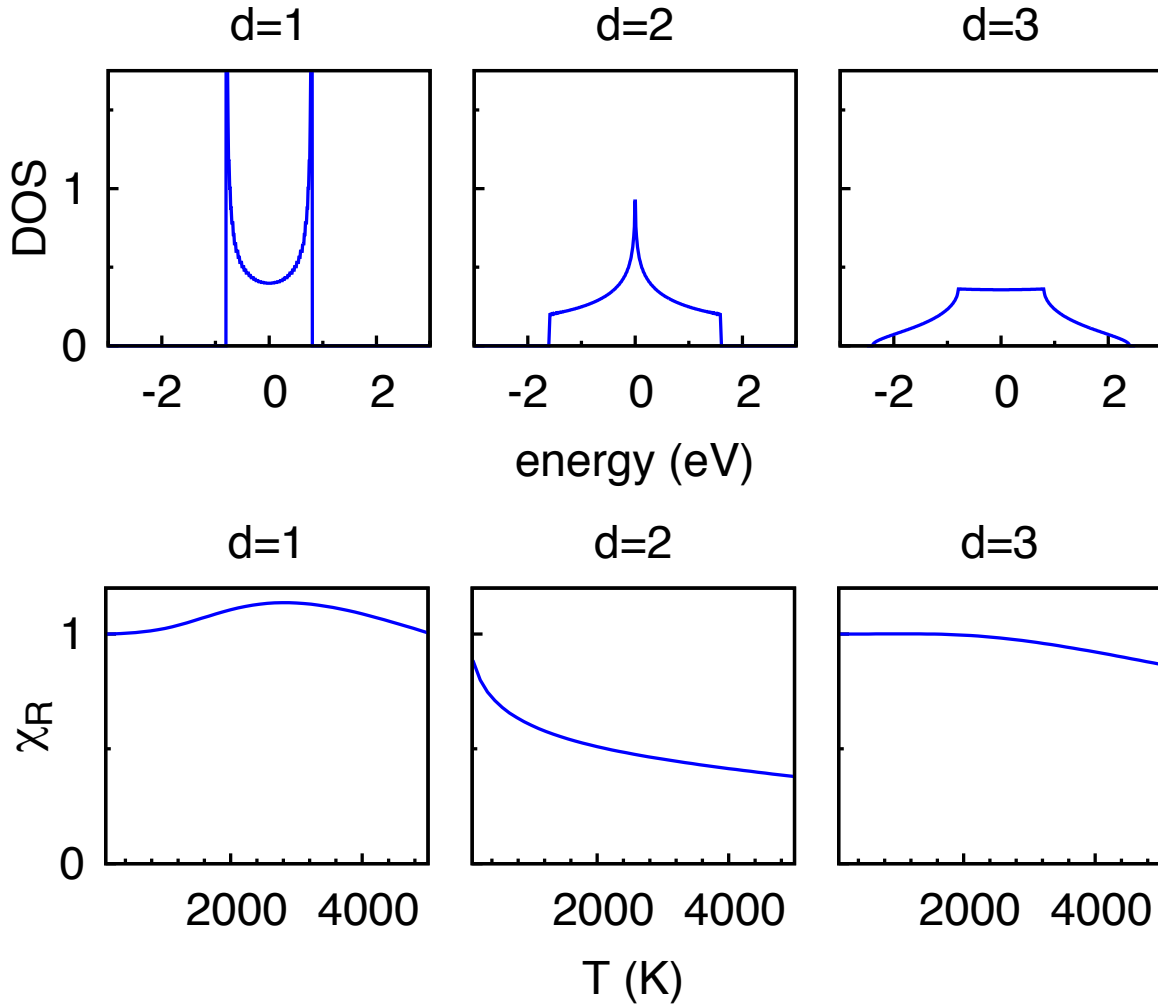


Fig. 3: *Top: Density of states (DOS) per spin, $\rho(\varepsilon)/2$, for a hypercubic lattice in one, two, and three dimension. The energy dispersion is calculated for $t = 0.4$ eV. The curves exhibit different types of Van-Hove singularities. Bottom: Effects of $\rho(\varepsilon_F)$ on the temperature dependence of $\chi_R = \chi^P(T)/\chi^P(0)$. Up to ~ 1000 K only the logarithmic Van-Hove singularity (two-dimensional case) yields a sizable effect.*

where $n_{\mathbf{k}\sigma} = c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma}$ and $N_{\mathbf{k}}$ is the number of \mathbf{k} points; $\rho(\varepsilon_F)$ is the total density of states (DOS) at the Fermi level, ε_F . The $T = 0$ susceptibility is then given by the Pauli formula

$$\chi^P(0) = \frac{1}{4} (g\mu_B)^2 \rho(\varepsilon_F).$$

In linear-response theory (see Appendix) the magnetization induced along \hat{z} by an external magnetic field $h_z(\mathbf{q}; \omega)\hat{z}$ oscillating with vector \mathbf{q} is given by

$$M_z(\mathbf{q}; \omega) = \chi_{zz}(\mathbf{q}; \omega) h_z(\mathbf{q}; \omega).$$

The Pauli susceptibility $\chi^P(0)$ is thus the static ($\omega = 0$) and uniform ($\mathbf{q} = 0$) linear response function to an external magnetic field. At finite temperature the Pauli susceptibility takes the

form

$$\chi^P(T) = \frac{1}{4} (g\mu_B)^2 \int d\varepsilon \rho(\varepsilon) \left(-\frac{dn(\varepsilon)}{d\varepsilon} \right),$$

where $n(\varepsilon) = 1/(1 + e^{(\varepsilon - \mu)\beta})$ is the Fermi distribution function, $\beta = 1/k_B T$ and μ the chemical potential. $\chi^P(T)$ depends weakly on the temperature; its temperature dependence is more pronounced, however, in the presence of van-Hove singularities close to the Fermi level (Fig. 3). Although we have considered here the non-interacting limit of the Hubbard model, Pauli paramagnetism is important even in the $U \neq 0$ case. This happens in the so-called *Fermi-liquid* regime. When Landau Fermi-liquid theory holds there is a one-to-one correspondence between the one-electron states and the excitations of the many-body system, the *quasi particles*. The latter are characterized by heavy masses m^*

$$\frac{m^*}{m} = 1 + \frac{1}{3} F_1^s > 1, \quad F_1^s > 0$$

and are more polarizable than electrons; correspondingly the system exhibits an enhanced Pauli susceptibility

$$\frac{\chi}{\chi^P} = \frac{1}{1 + F_0^a} > 1, \quad F_0^a < 0.$$

The coefficients F_1^s and F_0^a are Landau parameters. Because of the finite lifetime of quasiparticles and/or non Fermi-liquid phenomena of various nature, the temperature and energy regime in which the Fermi-liquid behavior is observed can be very narrow. This happens, e.g., for heavy Fermions or Kondo systems. We will discuss this in the last section.

2.1.2 Stoner instabilities

In the presence of the Coulomb interaction $U \neq 0$ finding the solution of the Hubbard model requires many-body techniques. Nevertheless, in the small U limit, we can already learn a lot about magnetism from Hartree-Fock (HF) static mean-field theory. In the simplest version of the HF approximation we make the following substitution

$$H_U = U \sum_i n_{i\uparrow} n_{i\downarrow} \rightarrow H_U^{\text{HF}} = U \sum_i [n_{i\uparrow} \langle n_{i\downarrow} \rangle + \langle n_{i\uparrow} \rangle n_{i\downarrow} - \langle n_{i\uparrow} \rangle \langle n_{i\downarrow} \rangle].$$

This approximation transforms the Coulomb two-particle interaction into an effective single-particle interaction. Let us search for a ferromagnetic solution and set therefore

$$\langle n_{i\sigma} \rangle = n_\sigma = \frac{n}{2} + \sigma m,$$

where $m = (n_\uparrow - n_\downarrow)/2$ and $n = n_\uparrow + n_\downarrow$. It is convenient to rewrite the mean-field Coulomb energy as in (10), i.e., as a function of m , n and S_z^i

$$H_U^{\text{HF}} = U \sum_i \left[-2m S_z^i + m^2 + \frac{n^2}{4} \right]. \quad (13)$$

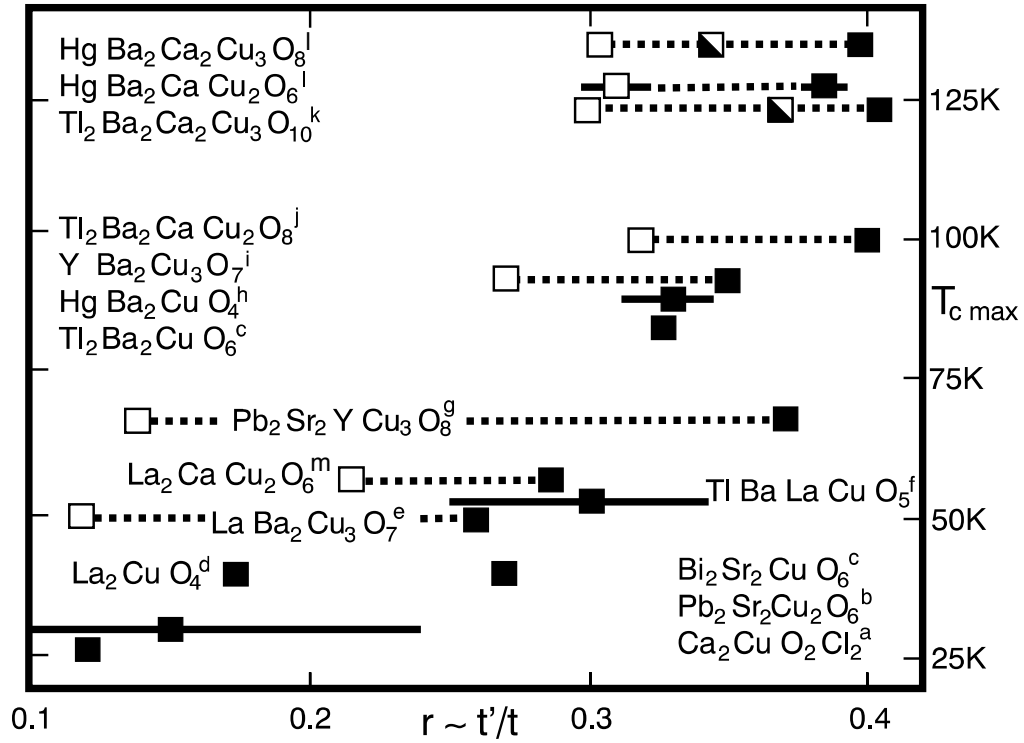


Fig. 4: Band-structure trend in hole-doped cuprates and correlation with $T_{c \text{ max}}$, the maximum value of the critical temperature for superconductivity. From Ref. [13].

The solution of the problem defined by the Hamiltonian $H_0 + H_U^{\text{HF}}$ amounts to the self-consistent solution of a non-interacting electron system with Bloch energies

$$\varepsilon_{\mathbf{k}\sigma}^U = \varepsilon_{\mathbf{k}} + n_{-\sigma} U = \varepsilon_{\mathbf{k}} + \frac{n}{2} U - \sigma m U.$$

In a magnetic field we additionally have to consider the Zeeman splitting. Thus

$$\varepsilon_{\mathbf{k}\sigma} = \varepsilon_{\mathbf{k}\sigma}^U + \frac{1}{2} g \mu_B h_z \sigma.$$

In the small U limit and for $T \rightarrow 0$ the magnetization $M_z = -g \mu_B m$ is then given by

$$M_z \sim \chi^P(0) \left[h_z - \frac{2}{g \mu_B} U m \right] = \chi^P(0) [h_z + 2(g \mu_B)^{-2} U M_z]$$

Solving for M_z we find the Stoner expression

$$\chi^S(\mathbf{0}; 0) = \frac{\chi^P(0)}{1 - 2(g \mu_B)^{-2} U \chi^P(0)}.$$

Thus with increasing U the $\mathbf{q} = \mathbf{0}$ static susceptibility increases and at the critical value

$$U_c = 2/\rho(\varepsilon_F)$$

it diverges, i.e., even an infinitesimal magnetic field can produce a finite magnetization. This means that the ground state becomes unstable against ferromagnetic order.

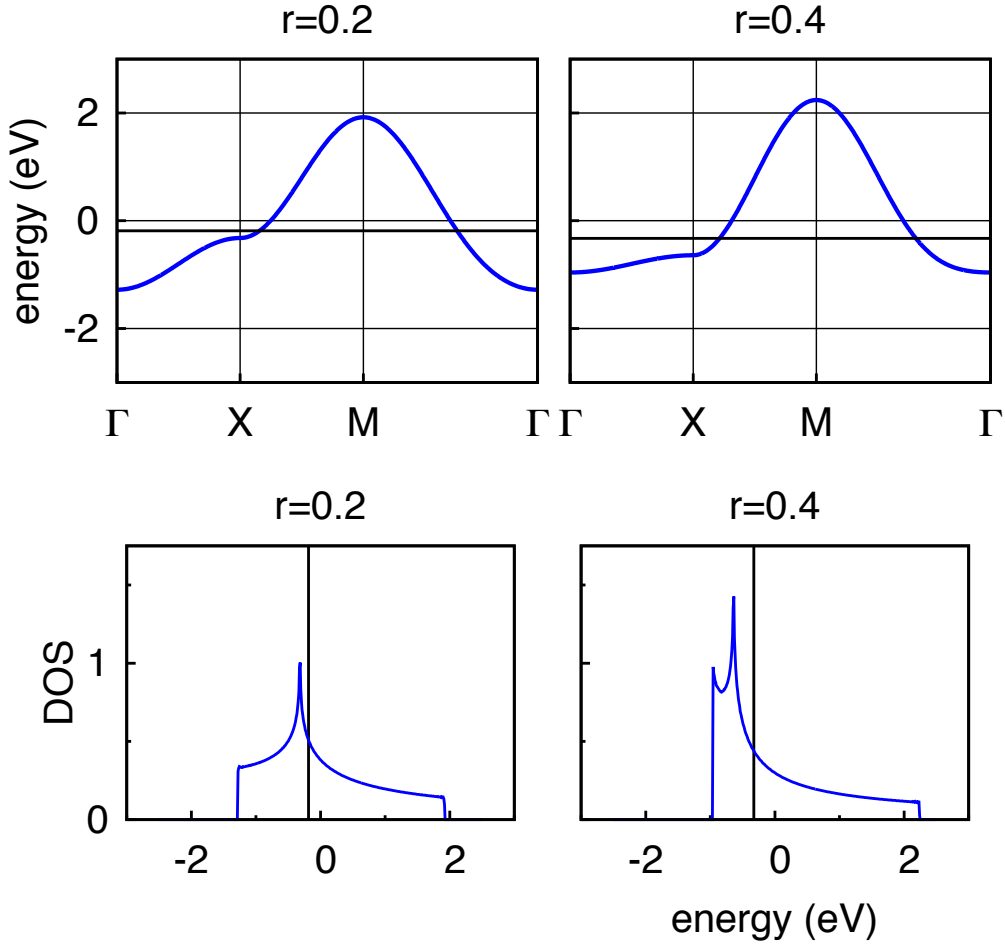


Fig. 5: *Top: Effect of $r = t'/t$ on the band structure of the two-dimensional tight-binding model. Black line: Fermi level at half-filling. Bottom: corresponding density of states per spin.*

Let us consider the case of the half-filled d -dimensional hypercubic lattice whose density of states is shown in Fig. 3. In three dimensions the DOS is flat around the Fermi level, e.g., $\rho(\varepsilon_F) \sim 2/W$ where W is the band width. For a flat DOS ferromagnetic instabilities are likely only when $U \sim W$, a rather large value of U , which typically also brings in strong-correlation effects not described by static mean-field theory. In two dimensions we have a rather different situation because a logarithmic Van-Hove singularity is exactly at the Fermi level (Fig. 3); a system with such a density of states is unstable toward ferromagnetism even for very small U . In real materials distortions or long-range interactions typically push the Van-Hove singularities away from the Fermi level. In HTSCs the electronic dispersion is modified as follows by the hopping t' between second nearest neighbors

$$\varepsilon_{\mathbf{k}} = -2t[\cos(k_x a) + \cos(k_y a)] + 4t' \cos(k_x a) \cos(k_y a).$$

As shown in Fig. 4, the parameter $r \sim t'/t$ ranges typically from ~ 0.15 to 0.4 [13]. Fig. 5 shows that with increasing r the Van-Hove singularity moves downwards in energy.

It is at this point natural to ask ourselves if ferromagnetism is the only possible instability. For a given system, magnetic instabilities with $\mathbf{q} \neq 0$ might be energetically favorable with respect

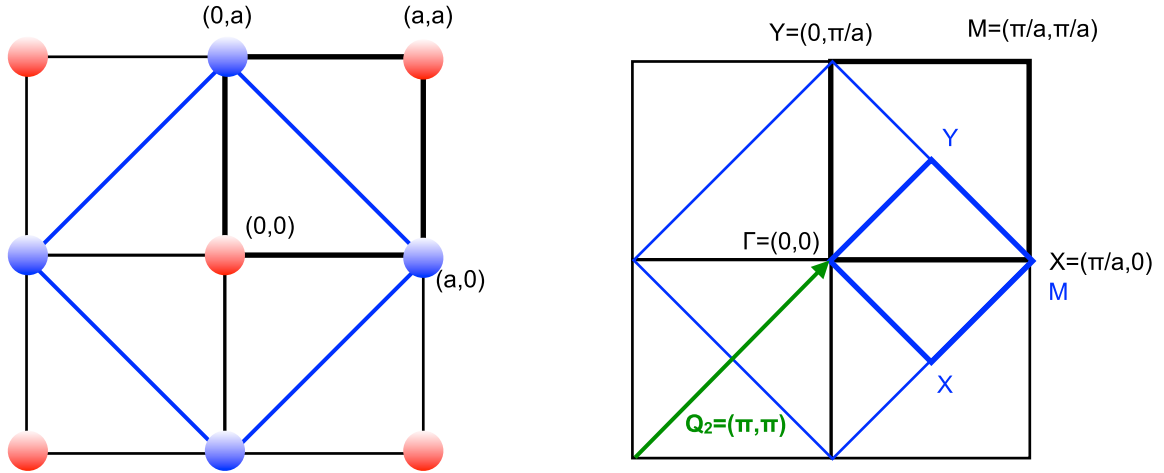


Fig. 6: Doubling of the cell due to antiferromagnetic order and corresponding folding of the Brillouin zone (BZ) for a two-dimensional hypercubic lattice. The antiferromagnetic $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$ vector is also shown.

to ferromagnetism; an example of a finite- \mathbf{q} instability is antiferromagnetism (see Fig. 6). To investigate finite- \mathbf{q} instabilities we generalize the Stoner criterion. Let us consider a magnetic excitation characterized by the vector \mathbf{q} commensurate with the reciprocal lattice. This magnetic superstructure defines a new lattice; the associated supercell includes $j = 1, \dots, N_j$ magnetically non-equivalent sites. We define therefore the quantities

$$S_z^i(\mathbf{q}) = \sum_j e^{i\mathbf{q} \cdot \mathbf{R}_j} S_z^{ji},$$

$$\langle S_z^{ji} \rangle = m \cos(\mathbf{q} \cdot \mathbf{R}_j),$$

where j runs over the magnetically non-equivalent sites $\{\mathbf{R}_j\}$ and i over the supercells in the lattice. In the presence of a magnetic field oscillating with vector \mathbf{q} and pointing in the z direction, $\mathbf{h}_j = h_z \cos(\mathbf{q} \cdot \mathbf{R}_j) \hat{z}$, the mean-field Coulomb and Zeeman terms can be written as

$$H_U^{\text{HF}} + H_Z = \sum_i \left[\frac{g\mu_B}{2} \left(h_z - \frac{2}{g\mu_B} mU \right) [S_z^i(\mathbf{q}) + S_z^i(-\mathbf{q})] + m^2 + \frac{n^2}{4} \right],$$

where m has to be determined self-consistently. This leads to the generalized Stoner formula

$$\chi^S(\mathbf{q}; 0) = \frac{1}{2} (g\mu_B)^2 \frac{\chi_0(\mathbf{q}; 0)}{[1 - U\chi_0(\mathbf{q}; 0)]}, \quad (14)$$

$$\chi_0(\mathbf{q}; 0) = -\frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \frac{n_{\mathbf{k}+\mathbf{q}} - n_{\mathbf{k}}}{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}.$$

The expression (14) is the same that we can find in the so-called random-phase approximation. For $\mathbf{q} = \mathbf{0}$ in the zero-temperature limit we recover the ferromagnetic RPA susceptibility with

$$\chi_0(\mathbf{0}; 0) = 2 (g\mu_B)^{-2} \chi^P(0) \sim \frac{1}{2} \rho(\varepsilon_F).$$

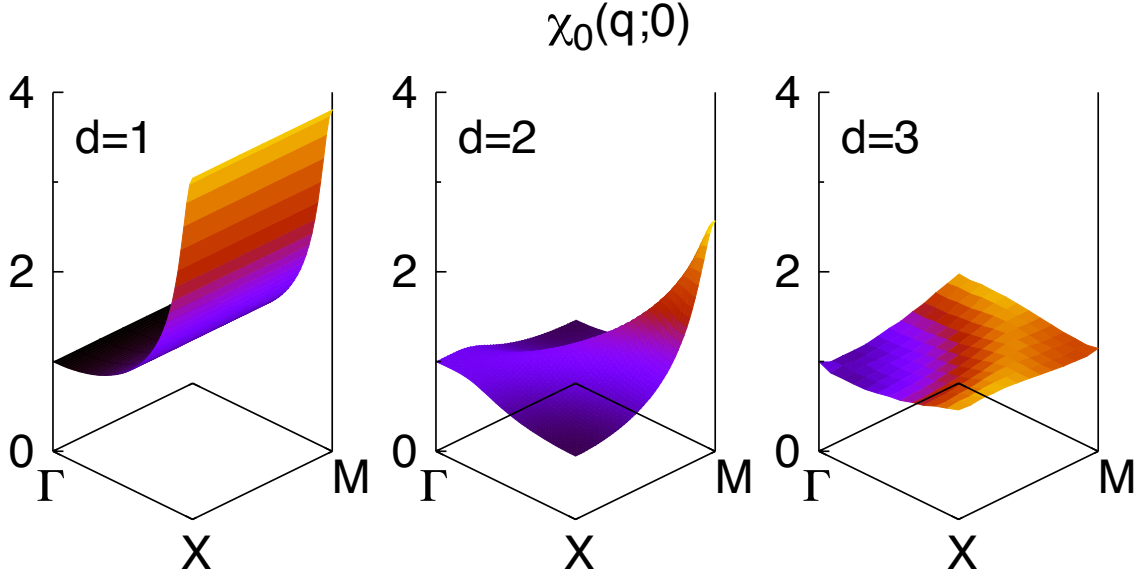


Fig. 7: The ratio $\chi_0(\mathbf{q};0)/\chi_0(\mathbf{0};0)$ in the xy plane for a hypercubic lattice ($T \sim 230$ K) at half filling. From left to right: one, two and three dimensions.

Figure 7 shows the non-interacting susceptibility in the xy plane for our d -dimensional hypercubic lattice. The figure shows that, in the one-dimensional case, the susceptibility diverges at the antiferromagnetic vector $\mathbf{Q}_1 = (\pi/a, 0, 0)$; in two dimensions this happens at $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$; in three dimension at $\mathbf{Q}_3 = (\pi/a, \pi/a, \pi/a)$, not shown in the figure. The energy dispersion exhibits at these vectors the property of perfect nesting

$$\varepsilon_{\mathbf{k}+\mathbf{Q}_i} = -\varepsilon_{\mathbf{k}}.$$

Remarkably, the $T = 0$ non-interacting susceptibility $\chi_0(\mathbf{Q}_i;0)$ diverges logarithmically at the nesting vector unless the density of states is zero at the Fermi level ($\varepsilon \rightarrow 0$)

$$\chi_0(\mathbf{Q}_i;0) \propto \frac{1}{4} \int_{-\infty}^{\varepsilon_F=0} d\varepsilon \rho(\varepsilon) \frac{1}{\varepsilon} \rightarrow \infty.$$

Under these conditions an arbitrary small U can cause a magnetic transition with magnetic vector \mathbf{Q}_i . In the two-dimensional case we have reached a similar conclusion for the $T = 0$ ferromagnetic ($\mathbf{q} = \mathbf{0}$) instability. The finite-temperature $\chi_0(\mathbf{q};0)$ susceptibility (Fig. 7) shows that, however, the antiferromagnetic instability is the strongest. Perfect nesting at \mathbf{Q}_2 is suppressed by $t' \neq 0$

$$\varepsilon_{\mathbf{k}+\mathbf{Q}_2} = -\varepsilon_{\mathbf{k}} + 8t' \cos(k_x a) \cos(k_y a).$$

Figure 8 shows how the susceptibility is modified by $t' \neq 0$ (half filling). The \mathbf{Q}_2 instability is important even for $t' \sim 0.4t$, but instabilities at incommensurate vectors around it are stronger. As a last remark it is important to notice that the RPA expression (14) depends on the filling only through the density of states, i.e., magnetic instabilities described by the Stoner formula can exist at *any* filling. This is very different from the case of the local moment regime that we will discuss starting from the next section.

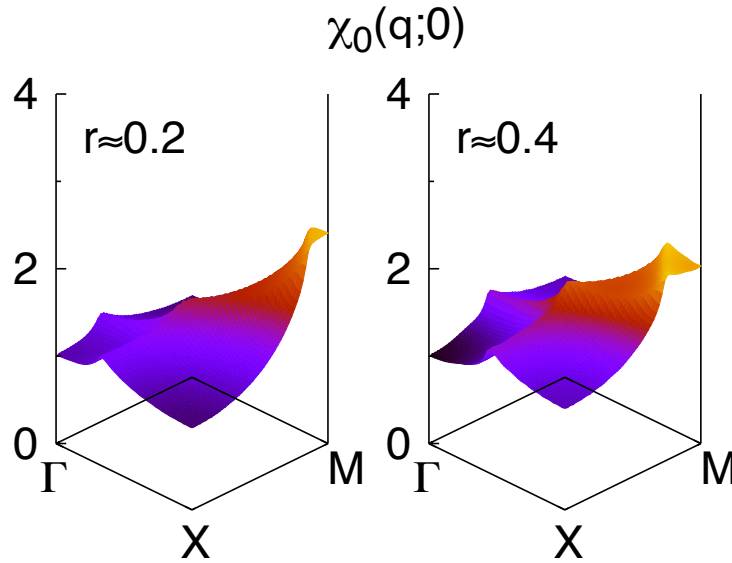


Fig. 8: The ratio $\chi_0(\mathbf{q};0)/\chi_0(\mathbf{0};0)$ in the xy plane for the two-dimensional hypercubic lattice (230 K) at half-filling. Left: $t' = 0.2t$. Right: $t' = 0.4t$.

2.2 Isolated magnetic ions

2.2.1 Paramagnetism

As we have seen, the ground-state multiplet of free ions with partially occupied shells can be determined via the Hund's rules. In Tab. 1 and Tab. 2 we can find the values of the S , L , and J quantum numbers for the ground-state multiplet of the most common transition-metal and rare-earth ions. If $t = 0$ and $n = 1$, the Hubbard model (7) describes precisely a collection of idealized free ions with an incomplete shell. For such idealized ions the only possible multiplet is the one with quantum numbers $J = S = 1/2, L = 0$. In the presence of a uniform external magnetic field $h_z \hat{z}$ we can then obtain the magnetization per atom as

$$M_z = \langle M_z^i \rangle = -g\mu_B \frac{\text{Tr} [e^{-g\mu_B h_z \beta S_z^i} S_z^i]}{\text{Tr} [e^{-g\mu_B h_z \beta S_z^i}]} = g\mu_B S \tanh(g\mu_B h_z \beta S),$$

and thus

$$\frac{\partial M_z}{\partial h_z} = (g\mu_B S)^2 \frac{1}{k_B T} [1 - \tanh^2(g\mu_B h_z \beta S)].$$

The static uniform susceptibility is then given by the $h \rightarrow 0$ limit

$$\chi_{zz}(\mathbf{0};0) = (g\mu_B S)^2 \frac{1}{k_B T} = \frac{C_{1/2}}{T}, \quad (15)$$

where $C_{1/2}$ is the $S = 1/2$ Curie constant. If $S = 1/2$, the relation $S^2 = S(S+1)/3$ holds. Thus, for reasons that will become clear in short, the Curie constant is typically expressed as

$$C_{1/2} = \frac{(g\mu_B)^2 S(S+1)}{3k_B}.$$

Ion		n	S	L	J	$2S+1L_J$
V ⁴⁺	Ti ³⁺	$3d^1$	1/2	2	3/2	$^2D_{3/2}$
	V ³⁺	$3d^2$	1	3	2	2F_2
	Cr ³⁺ V ²⁺	$3d^3$	3/2	3	3/2	$^4F_{3/2}$
	Mn ³⁺ Cr ²⁺	$3d^4$	2	2	0	5D_0
	Fe ³⁺ Mn ²⁺	$3d^5$	5/2	0	5/2	$^6S_{5/2}$
	Fe ²⁺	$3d^6$	2	2	4	5D_4
	Co ²⁺	$3d^7$	3/2	3	9/2	$^4F_{9/2}$
	Ni ²⁺	$3d^8$	1	3	4	3F_4
	Cu ²⁺	$3d^9$	1/2	2	5/2	$^2D_{5/2}$

Table 1: Quantum numbers of the ground-state multiplet for several transition-metal ions with partially filled d shells. In transition-metal oxides the angular momentum is typically quenched because of the crystal-field and therefore only the total spin matters.

If the ions have ground-state total angular momentum J we can calculate the susceptibility with the same technique, provided that we replace g with the Landé factor g_J

$$g_J = \frac{\langle JJ_zLS|(g\mathbf{S} + \mathbf{L}) \cdot \mathbf{J}|JJ_zLS\rangle}{\langle JJ_zLS|\mathbf{J} \cdot \mathbf{J}|JJ_zLS\rangle}$$

$$\sim \frac{3}{2} + \frac{S(S+1) - L(L+1)}{2J(J+1)},$$

and calculate the thermal average of the magnetization, $\mathbf{M} = -g_J\mu_B\mathbf{J}$, accounting for the $2J+1$ degeneracy of the multiplet. The result is

$$M_z = \langle M_z^i \rangle = g_J\mu_B J B_J(g_J\mu_B h_z \beta J)$$

where $B_J(x)$ is the Brillouin function

$$B_J(x) = \frac{2J+1}{2J} \coth\left(\frac{2J+1}{2J}x\right) - \frac{1}{2J} \coth\left(\frac{1}{2J}x\right).$$

In the low-temperature ($x \rightarrow \infty$) limit $B_J(x) \sim 1$, and thus the magnetization approaches its saturation value in which all atoms are in the ground state

$$M_z \sim g_J\mu_B J \equiv M_0.$$

In the high-temperature ($x \rightarrow 0$) limit

$$B_J(x) \sim x \frac{J+1}{3J} \left[1 - \frac{2J^2 + 2J + 1}{30J^2} x^2 \right],$$

and thus the susceptibility exhibits the Curie high-temperature behavior

$$\chi_{zz}(\mathbf{0}; 0) \sim \frac{C_J}{T} = \frac{\mu^2}{3k_B T},$$

Ion	n	S	L	J	$^{2S+1}L_J$	g_J
Ce ³⁺	4f ¹	1/2	3	5/2	² F _{5/2}	6/7
Pr ³⁺	4f ²	1	5	4	³ H ₄	4/5
Nd ³⁺	4f ³	3/2	6	9/2	⁴ I _{9/2}	8/11
Pm ³⁺	4f ⁴	2	6	4	⁵ I ₄	3/5
Sm ³⁺	4f ⁵	5/2	5	5/2	⁶ H _{5/2}	2/7
Eu ³⁺	4f ⁶	3	3	0	⁷ F ₀	0
Gd ³⁺	4f ⁷	7/2	0	7/2	⁸ S _{7/2}	2
Tb ³⁺	4f ⁸	3	3	6	⁷ F ₆	3/2
Dy ³⁺	4f ⁹	5/2	5	15/2	⁶ H _{15/2}	4/3
Ho ³⁺	4f ¹⁰	2	6	8	⁵ I ₈	5/4
Er ³⁺	4f ¹¹	3/2	6	15/2	⁴ I _{15/2}	6/5
Tm ³⁺	4f ¹²	1	5	6	³ H ₆	7/6
Yb ³⁺	4f ¹³	1/2	3	7/2	² F _{7/2}	8/7

Table 2: Quantum numbers of the ground-state multiplet for rare-earth ions with partially filled f shells and corresponding g_J factor. In $4f$ materials the crystal field is typically small; thus the ground-state multiplet is in first approximation close to that of the corresponding free ion.

where the generalized Curie constant is

$$C_J = \frac{(g_J \mu_B)^2 J(J+1)}{3k_B},$$

and where $\mu = g_J \mu_B \sqrt{J(J+1)}$ is the total magnetic moment. Correspondingly, the susceptibility decreases as $1/T$ with increasing T (Fig. 9). We have thus the three limit cases

$$\chi_{zz}(\mathbf{0}; 0) \sim \begin{cases} 0 & k_B T / |M_0| h_z \rightarrow 0 \\ C_J / T & |M_0| h_z / k_B T \rightarrow 0 \\ C_J / T & h_z \rightarrow 0 \end{cases}.$$

Remarkably, the $T \rightarrow 0$ and $h_z \rightarrow 0$ limit cannot be interchanged. If h_z is finite the susceptibility goes to zero in the $T \rightarrow 0$ limit; instead, if we perform the $h_z \rightarrow 0$ limit first it diverges with the Curie form $1/T$. The point $h_z = T = 0$ is a critical point in the phase space.

Let us return to the $S = 1/2$ case, i.e., the one relevant for the Hubbard model. It is interesting to calculate the inter-site spin correlation function $\mathcal{S}_{i,i'}$

$$\mathcal{S}_{i,i'} = \langle (\mathbf{S}_i - \langle \mathbf{S}_i \rangle) \cdot (\mathbf{S}_{i'} - \langle \mathbf{S}_{i'} \rangle) \rangle = \langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle - \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle.$$

We express $\langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle$ in the form $[S(S+1) - S_i(S_i+1) - S_{i'}(S_{i'}+1)]/2$, where $S_i = S_{i'} = 1/2$ and $\mathbf{S} = \mathbf{S}_i + \mathbf{S}_{i'}$ is the total spin. Then, since in the absence of magnetic field $\langle \mathbf{S}_i \rangle = \langle \mathbf{S}_{i'} \rangle = 0$,

$$\mathcal{S}_{i,i'} = [S(S+1) - 3/2]/2 = \begin{cases} 1/4 & S = 1 \\ -3/4 & S = 0 \end{cases}.$$

The ideal paramagnetic state is however characterized by uncorrelated sites. Hence

$$\mathcal{S}_{i,i'} = \langle \mathbf{S}_i \cdot \mathbf{S}_{i'} \rangle \sim \begin{cases} \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle & \sim 0 & i \neq i' \\ \langle \mathbf{S}_i \cdot \mathbf{S}_i \rangle & = 3/4 & i = i' \end{cases}. \quad (16)$$

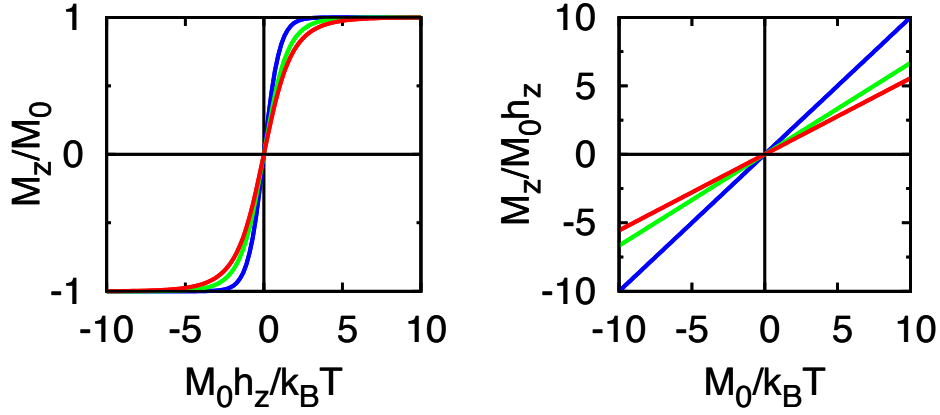


Fig. 9: Left: $M_z/M_0 = B_J(x)$ as a function of $x = h_z M_0 / k_B T$. The different lines correspond to $J = 1/2$ (blue), $J = 1$ (green) and $J = 3/2$ (red). Right: The ratio $M_z/M_0 h_z$ for finite magnetic field in the small x limit; the slope is $(J + 1)/3J$.

The (ideal) paramagnetic phase is thus quite different from a spatially disordered state, i.e., a situation in which each ion has a spin oriented in a given direction but spin orientations are randomly distributed. In the latter case, in general, $\langle \mathbf{S}^i \cdot \mathbf{S}^{i'} \rangle \neq 0$ for $i' \neq i$, even if, e.g., the sum of $\langle S_z^i \cdot S_z^{i'} \rangle$ over all sites i' with $i' \neq i$ is zero

$$\sum_{i' \neq i} \langle S_z^i \cdot S_z^{i'} \rangle \sim 0.$$

The high-temperature static susceptibility can be obtained from the correlation function Eq. (16) using the *fluctuation-dissipation theorem* and the Kramers-Kronig relations (see Appendix). The result is

$$\chi_{zz}(\mathbf{q}; 0) \sim \frac{(g\mu_B)^2}{k_B T} \sum_{i'} \mathcal{S}_{zz}^{i,i'} e^{i\mathbf{q} \cdot (\mathbf{R}_i - \mathbf{R}_{i'})} = \chi_{zz}^i(T) = \frac{M_0^2}{k_B T} = \frac{C_{1/2}}{T}. \quad (17)$$

This shows that $\chi_{zz}(\mathbf{q}; 0)$ is \mathbf{q} -independent and coincides with the local susceptibility $\chi_{zz}^i(T)$

$$\chi_{zz}(\mathbf{0}; 0) = \lim_{h_z \rightarrow 0} \frac{\partial M_z}{\partial h_z} = \chi_{zz}^i(T).$$

How can the spin susceptibility (17) be obtained directly from the atomic limit of the Hubbard model, Eq. (10)? To calculate it we can use, e.g., the imaginary time and Matsubara frequencies formalism (see Appendix). Alternatively at high temperatures we can obtain it from the correlation function as we have just seen. The energy of the four atomic states are given by (9) and, at half filling, the chemical potential is $\mu = \varepsilon_d + U/2$. Therefore

$$\begin{aligned} \chi_{zz}(\mathbf{0}; 0) &\sim \frac{(g\mu_B)^2}{k_B T} \left\{ \frac{\text{Tr} \left[e^{-\beta(H_i - \mu N_i)} (S_z^i)^2 \right]}{\text{Tr} \left[e^{-\beta(H_i - \mu N_i)} \right]} - \left[\frac{\text{Tr} \left[e^{-\beta(H_i - \mu N_i)} S_z^i \right]}{\text{Tr} \left[e^{-\beta(H_i - \mu N_i)} \right]} \right]^2 \right\} \\ &= \frac{C_{1/2}}{T} \frac{e^{\beta U/2}}{1 + e^{\beta U/2}}. \end{aligned}$$

Thus the susceptibility depends on the energy scale

$$U = E(N_i + 1) + E(N_i - 1) - 2E(N_i).$$

If we perform the limit $U \rightarrow \infty$ we effectively eliminate doubly occupied and empty states. In this limit we recover the expression that we found for the spin $S = 1/2$ model, Eq. (17). This is a trivial example of downfolding, in which the low-energy and high-energy sector are decoupled from the start in the Hamiltonian. In the large U limit the high-energy states are integrated out leaving the system in a magnetic $S = 1/2$ state.

2.2.2 Larmor diamagnetism and Van Vleck paramagnetism

For ions with $J = 0$ the ground-state multiplet, in short $|0\rangle$, is non-degenerate and the linear correction to the ground-state total energy due to the Zeeman term is zero; remarkably, for open-shell ions the magnetization remains nevertheless finite because of higher-order corrections. At second order there are two contributions for the ground state. The first is the Van-Vleck term

$$M_z^{\text{VV}} = 2h_z\mu_B^2 \sum_I \frac{|\langle 0|(L_z + gS_z)|I\rangle|^2}{E_I - E_0},$$

where E_I is the energy of the excited state $|I\rangle$ and E_0 the energy of the ground-state multiplet. The Van-Vleck term is weakly temperature-dependent and typically small. The second term is the diamagnetic Larmor contribution

$$M_z^{\text{L}} = -\frac{1}{4}h_z\langle 0|\sum_i(x_i^2 + y_i^2)|0\rangle.$$

The Larmor and Van-Vleck terms have opposite signs and typically compete with each other.

2.3 Interacting localized moments

2.3.1 Spin models

In the large U limit and at half filling we can map the Hubbard model into an effective Heisenberg model. In this section we solve the latter using static mean-field theory. In the mean-field approximation we replace the Heisenberg Hamiltonian (11) with

$$H_S^{\text{MF}} = \frac{1}{2}\Gamma \sum_{\langle ii'\rangle} \left[\mathbf{S}_i \cdot \langle \mathbf{S}_{i'} \rangle + \langle \mathbf{S}_i \rangle \cdot \mathbf{S}_{i'} - \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_{i'} \rangle - \frac{1}{4}n_i n_{i'} \right].$$

In the presence of an external magnetic field \mathbf{h} we add the Zeeman term and have in total

$$\begin{aligned} H &= g\mu_B \sum_i [\mathbf{S}_i \cdot (\mathbf{h} + \mathbf{h}_i^m) + \text{const}] , \\ \mathbf{h}_i^m &= n_{\langle ii'\rangle} \Gamma \langle \mathbf{S}_{i'} \rangle / g\mu_B , \end{aligned}$$

where $n_{\langle ii'\rangle}$ is the number of first nearest neighbors and \mathbf{h}_i^m is the molecular field at site i . We define the quantization axis z as the direction of the external magnetic field, $\mathbf{h} = h_z \hat{z}$,

and assume that \hat{z} is also the direction of the molecular field, $\mathbf{h}_i^m = \Delta h_z^i \hat{z}$. Since $\Gamma > 0$ and hypercubic lattices are bipartite, the likely magnetic order is two-sublattice antiferromagnetism. Thus we set $M_z^A = -g\mu_B \langle S_z^i \rangle$, $M_z^B = -g\mu_B \langle S_z^{i'} \rangle$, where A and B are the two sublattices, $i \in A$ and $i' \in B$. In the absence of an external magnetic field, the total magnetization per formula unit, $M_z = (M_z^B + M_z^A)/2$, vanishes in the antiferromagnetic state. We define therefore as order parameter $\sigma_m = 2m = (M_z^B - M_z^A)/2M_0$, which is zero only above the critical temperature for antiferromagnetic order. We then calculate the magnetization for each sublattice and find the system of coupled equations

$$\begin{cases} M_z^A/M_0 &= B_{1/2} [M_0(h_z + \Delta h_z^A)\beta] \\ M_z^B/M_0 &= B_{1/2} [M_0(h_z + \Delta h_z^B)\beta] \end{cases}, \quad (18)$$

where

$$\begin{cases} \Delta h_z^A &= -(M_z^B/M_0) S^2 \Gamma n_{\langle ii' \rangle} / M_0 \\ \Delta h_z^B &= -(M_z^A/M_0) S^2 \Gamma n_{\langle ii' \rangle} / M_0 \end{cases}.$$

For $h_z = 0$ the system (18) can be reduced to the single equation

$$\sigma_m = B_{1/2} [\sigma_m S^2 \Gamma n_{\langle ii' \rangle} \beta]. \quad (19)$$

This equation has always the trivial solution $\sigma_m = 0$. Figure 10 shows that, for small enough temperatures it also has a non-trivial solution $\sigma_m \neq 0$. The order parameter σ_m equals ± 1 at zero temperature and its absolute value decreases with increasing temperature. It becomes zero for $T \geq T_N$ with

$$k_B T_N = \frac{S(S+1)}{3} n_{\langle ii' \rangle} \Gamma.$$

If $T \sim T_N$ we can find the non-trivial solution by first rewriting (19) as

$$\sigma_m = B_{1/2} \left[\frac{T_N}{T} \sigma_m \right]. \quad (20)$$

The inverse of this equation yields T/T_N as a function of σ_m

$$\frac{T}{T_N} = \frac{\sigma_m}{B_{1/2}^{-1}[\sigma_m]}.$$

If $T \sim T_N$ the parameter σ_m is small. We then expand the right-hand side in powers of σ_m

$$\frac{\sigma_m}{B_{1/2}^{-1}(\sigma_m)} \sim \frac{\sigma_m}{\sigma_m + \sigma_m^3/3 + \dots} \sim 1 - \sigma_m^2/3 + \dots$$

This leads to the following expression

$$\sigma_m = \sqrt{3} \left(1 - \frac{T}{T_N} \right)^{1/2},$$

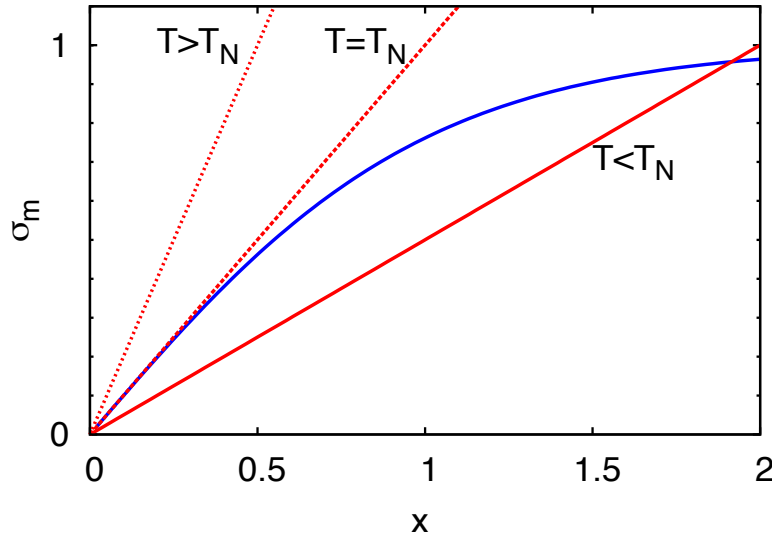


Fig. 10: The self-consistent solution of Eq. (20) for $\sigma_m \geq 0$. The blue line shows the right-hand side of the equation, the Brillouin function $B_{1/2}(x)$, with $x = \sigma_m T_N / T$. The red lines show the left-hand side of the equation, $\sigma_m(x) = \alpha x$, with $\alpha = T / T_N$; the three different curves correspond to representative T / T_N values.

which shows that the order parameter has a discontinuous temperature derivative at $T = T_N$.

It is interesting to derive the expression of the static uniform susceptibility. For this we go back to the system of equations (18) and calculate from it the total magnetization M_z . In the weak magnetic field limit $M_z^A \sim -\sigma_m M_0 + \chi_{zz}(\mathbf{0}; 0) h_z$ and $M_z^B \sim \sigma_m M_0 + \chi_{zz}(\mathbf{0}; 0) h_z$. Then, by performing the first derivative of M_z with respect to h_z in the $h_z \rightarrow 0$ limit we obtain

$$\chi_{zz}(\mathbf{0}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T + (1 - \sigma_m^2)T_N}.$$

The uniform susceptibility vanishes at $T = 0$ and reaches the maximum at $T = T_N$, where it takes the value $C_{1/2}/2T_N$. In the high-temperature regime $\sigma_m = 0$ and

$$\chi_{zz}(\mathbf{0}; 0) \sim \frac{C_{1/2}}{T + T_N},$$

which is *smaller* than the susceptibility of free $S = 1/2$ magnetic ions.

The magnetic linear response is quite different if we apply an external field \mathbf{h}_\perp perpendicular to the spins in the antiferromagnetic lattice. The associated perpendicular magnetization is

$$M_\perp \sim M_0 \frac{\sigma_m (g\mu_B h_\perp)}{\sqrt{(g\mu_B h_\perp)^2 + (4\sigma_m)^2 (k_B T_N)^2}},$$

and therefore the perpendicular susceptibility is temperature-independent for $T \leq T_N$

$$\chi_\perp(\mathbf{0}; 0) = \lim_{h_\perp \rightarrow 0} \frac{dM_\perp}{dh_\perp} = \frac{C_{1/2}}{2T_N}.$$

Hence, for $T < T_N$ the susceptibility is anisotropic, $\chi_{zz}(\mathbf{0}; 0) = \chi_{\parallel}(\mathbf{0}; 0) \neq \chi_\perp(\mathbf{0}; 0)$; at the absolute zero $\chi_{\parallel}(\mathbf{0}; 0)$ vanishes, but the response to \mathbf{h}_\perp remains strong. For $T > T_N$ the order parameter is zero and the susceptibility isotropic, $\chi_{\parallel}(\mathbf{0}; 0) = \chi_\perp(\mathbf{0}; 0)$.

We have up to now considered antiferromagnetic order only. What about other magnetic instabilities? Let us consider first ferromagnetic order. For a ferromagnetic spin arrangement by repeating the calculation we find

$$\chi_{zz}(\mathbf{0}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T - (1 - \sigma_m^2)T_C},$$

where $T_C = -S(S+1)n_{\langle ii' \rangle} \Gamma / 3k_B$ is, if the exchange coupling Γ is negative, the critical temperature for ferromagnetic order. Then, differently than in the antiferromagnetic case, the high-temperature uniform susceptibility is *larger* than that of free $S = 1/2$ magnetic ions.

For a generic magnetic structure characterized by a vector \mathbf{q} and a supercell with $j = 1, \dots, N_j$ magnetically non-equivalent sites we make the Ansatz

$$\langle M_z^{ji} \rangle = -\sigma_m M_0 \cos(\mathbf{q} \cdot \mathbf{R}_j) = -g\mu_B m \cos(\mathbf{q} \cdot \mathbf{R}_j),$$

where σ_m is again the order parameter. We consider a magnetic field rotating with the same \mathbf{q} vector. By using the static mean-field approach we then find

$$k_B T_{\mathbf{q}} = \frac{S(S+1)}{3} \Gamma_{\mathbf{q}}, \quad \Gamma_{\mathbf{q}} = - \sum_{ij \neq 0} \Gamma^{00,ij} e^{i\mathbf{q} \cdot (\mathbf{T}_i + \mathbf{R}_j)}, \quad (21)$$

where $\Gamma^{00,ij}$ is the exchange coupling between the spin at the origin and the spin at site ij , and $\{\mathbf{T}_i\}$ are lattice vectors. In our example, $T_0 = T_C$ and $T_{\mathbf{q}_{\text{AF}}} = T_N = -T_C$. Thus we have

$$\chi_{zz}(\mathbf{q}; 0) = \frac{C_{1/2}(1 - \sigma_m^2)}{T - (1 - \sigma_m^2)T_{\mathbf{q}}}, \quad (22)$$

which diverges at $T = T_{\mathbf{q}}$. The susceptibility $\chi_{zz}(\mathbf{q}; 0)$ reflects the spatial extension of correlations, i.e., the *correlation length*, ξ ; the divergence of the susceptibility at $T_{\mathbf{q}}$ is closely related to the divergence of ξ . To see this we calculate ξ for a hypercubic three-dimensional lattice, assuming that the system has only one instability with vector \mathbf{Q} . First we expand Eq. (21) around \mathbf{Q} obtaining $T_{\mathbf{q}} \sim T_{\mathbf{Q}} + \alpha(\mathbf{q} - \mathbf{Q})^2 + \dots$ and then we calculate $\chi_{zz}^{00,ji}$, the Fourier transform of Eq. (22). We find that $\chi_{zz}^{00,ji}$ decays exponentially with $r = |\mathbf{T}_i + \mathbf{R}_j|$, i.e., $\chi_{zz}^{00,ji} \propto e^{-r/\xi}/r$. The range of the correlations is $\xi \propto [T_{\mathbf{Q}}/(T - T_{\mathbf{Q}})]^{1/2}$, which becomes infinite at $T = T_{\mathbf{Q}}$.

It is important to notice that in principle there can be instabilities at any \mathbf{q} vector, i.e., \mathbf{q} does not need to be commensurate with reciprocal lattice vectors. The value of \mathbf{q} for which $T_{\mathbf{q}}$ is the largest determines (within static mean-field theory) the type of magnetic order that is realized. The antiferromagnetic structure in Fig. 6 corresponds to $\mathbf{q}_{\text{AF}} = \mathbf{Q}_2 = (\pi/a, \pi/a, 0)$.

In real systems the spin S is typically replaced by an *effective magnetic moment*, μ_{eff} , and therefore $C_{1/2} \rightarrow C_{\text{eff}} = \mu_{\text{eff}}^2 / 3k_B$. It follows that μ_{eff} is the value of the product $3k_B T \chi_{zz}(\mathbf{q}; 0)$ in the high-temperature limit (here $T \gg T_{\mathbf{q}}$). The actual value of μ_{eff} depends, as we have discussed in the introduction, on the Coulomb interaction, the spin-orbit coupling and the crystal field. In addition, the effective moment can be screened by many-body effects, as it happens for Kondo impurities; we will discuss the latter case in the last section.

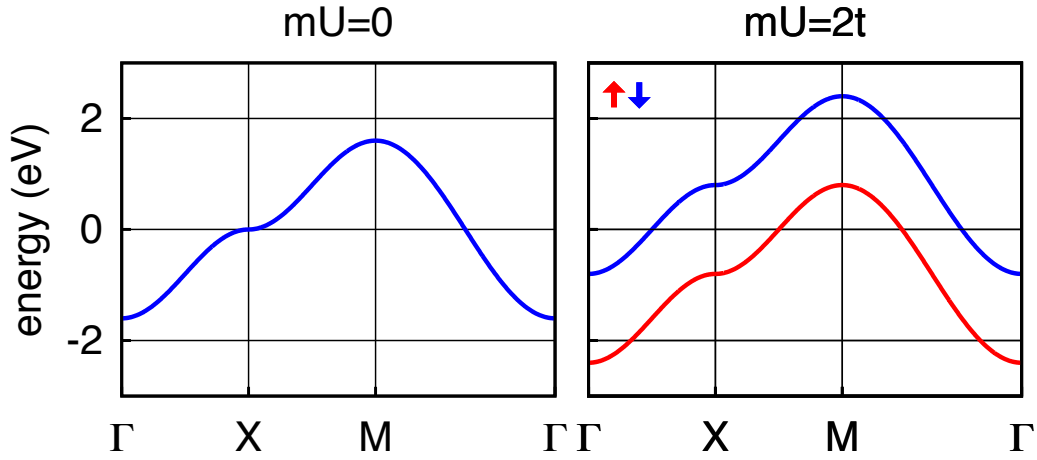


Fig. 11: *Ferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero.*

2.3.2 The Hartree-Fock approximation

We have seen that Hartree-Fock mean-field theory yields Stoner magnetic instabilities in the weak coupling limit. Can it also describe magnetism in the local moment regime ($t/U \ll 1$)? Let us focus on the half-filled two-dimensional Hubbard model for a square lattice, and let us analyze two possible magnetically ordered states, the ferro- and the antiferro-magnetic state.

If we are only interested in the ferromagnetic or the paramagnetic solution, the Hartree-Fock approximation of the Coulomb term in the Hubbard model, H_U^{HF} , is given by Eq. (13); the HF Hamiltonian is $H = H_d + H_T + H_U^{\text{HF}}$. For periodic systems it is convenient to write H in \mathbf{k} space. We then adopt as one-electron basis the Bloch states $\Psi_{\mathbf{k}\sigma}$

$$\Psi_{\mathbf{k}\sigma}(\mathbf{r}) = \frac{1}{\sqrt{N_s}} \sum_i e^{i\mathbf{k} \cdot \mathbf{T}_i} \Psi_{i\sigma}(\mathbf{r}),$$

where $\Psi_{i\sigma}(\mathbf{r})$ is a Wannier function with spin σ , \mathbf{T}_i a lattice vector and N_s the number of lattice sites. The term H_U^{HF} depends on the spin operator S_z^i , whose Fourier transform in \mathbf{k} space is

$$S_z(\mathbf{k}, \mathbf{k}') = \frac{1}{N_s} \sum_i e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{T}_i} \frac{1}{2} \sum_{\sigma} \sigma c_{i\sigma}^{\dagger} c_{i\sigma}.$$

The term H_U^{HF} has the same periodicity of the lattice and does not couple states with different \mathbf{k} vectors. Thus only $S_z(\mathbf{k}, \mathbf{k})$ contributes, and the Hamiltonian can be written as

$$H = \sum_{\sigma} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + U \sum_{\mathbf{k}} \left[-2m S_z(\mathbf{k}, \mathbf{k}) + m^2 + \frac{n^2}{4} \right],$$

where $m = (n_{\uparrow} - n_{\downarrow})/2$ and $n = 1$; for simplicity we set $\varepsilon_d = 0$. The HF correction splits the bands with opposite spin, leading to new one-electron eigenvalues, $\varepsilon_{\mathbf{k}\sigma} = \varepsilon_{\mathbf{k}} + \frac{1}{2}U - \sigma Um$; the chemical potential is $\mu = U/2$. The separation between $\varepsilon_{\mathbf{k}\uparrow} - \mu$ and $\varepsilon_{\mathbf{k}\downarrow} - \mu$ is $2mU$, as can be seen in Fig. 11. The system remains metallic for U smaller than the bandwidth W . In the

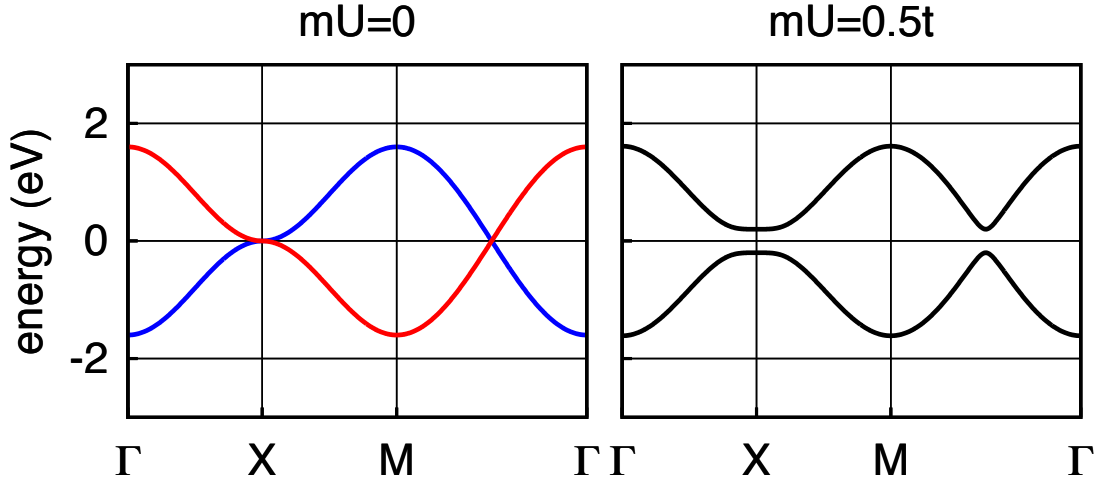


Fig. 12: Antiferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero. Blue: $\varepsilon_{\mathbf{k}}$. Red: $\varepsilon_{\mathbf{k}+\mathbf{Q}_2} = -\varepsilon_{\mathbf{k}}$. The high-symmetry lines are those of the large BZ in Fig. 6.

large t/U limit and at half filling we can assume that the system is a ferromagnetic insulator and $m = 1/2$. The total energy of the ground state is then

$$E_F = \frac{1}{N_k} \sum_{\mathbf{k}} [\varepsilon_{\mathbf{k}\sigma} - \mu] = \frac{1}{N_k} \sum_{\mathbf{k}} \left[\varepsilon_{\mathbf{k}} - \frac{1}{2}U \right] = -\frac{1}{2}U.$$

Let us now describe the same periodic lattice via a supercell which allows for a two-sublattice antiferromagnetic solution; this supercell is shown in Fig. 6. We rewrite the Bloch states of the original lattice as

$$\Psi_{\mathbf{k}\sigma}(\mathbf{r}) = \frac{1}{\sqrt{2}} [\Psi_{\mathbf{k}\sigma}^A(\mathbf{r}) + \Psi_{\mathbf{k}\sigma}^B(\mathbf{r})], \quad \Psi_{\mathbf{k}\sigma}^\alpha(\mathbf{r}) = \frac{1}{\sqrt{N_{s_\alpha}}} \sum_{i_\alpha} e^{i\mathbf{T}_{i_\alpha}^\alpha \cdot \mathbf{k}} \Psi_{i_\alpha\sigma}(\mathbf{r}).$$

Here A and B are the two sublattices with opposite spins and \mathbf{T}_i^A and \mathbf{T}_i^B are their lattice vectors; $\alpha = A, B$. We take as one-electron basis the two Bloch functions $\Psi_{\mathbf{k}\sigma}$ and $\Psi_{\mathbf{k}+\mathbf{Q}_2\sigma}$, where $\mathbf{Q}_2 = (\pi/a, \pi/a, 0)$ is the vector associated with the antiferromagnetic instability and the corresponding folding of the Brillouin zone, also shown in Fig. 6. Then, in the HF approximation, the Coulomb interaction is given by

$$H_U^{\text{HF}} = \sum_{i \in A} \left[-2mS_z^i + m^2 + \frac{n^2}{4} \right] + \sum_{i \in B} \left[+2mS_z^i + m^2 + \frac{n^2}{4} \right].$$

This interaction couples Bloch states with \mathbf{k} vectors made equivalent by the folding of the Brillouin zone. Thus the HF Hamiltonian takes the form

$$H = \sum_{\mathbf{k}} \sum_{\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{\mathbf{k}} \sum_{\sigma} \varepsilon_{\mathbf{k}+\mathbf{Q}_2} n_{\mathbf{k}+\mathbf{Q}_2\sigma} + U \sum_{\mathbf{k}} \left[-2m S_z(\mathbf{k}, \mathbf{k} + \mathbf{Q}_2) + 2m^2 + 2\frac{n^2}{4} \right].$$

The sum over \mathbf{k} is restricted to the Brillouin zone of the antiferromagnetic lattice. We find the two-fold degenerate eigenvalues

$$\varepsilon_{\mathbf{k}\pm} - \mu = \frac{1}{2}(\varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}+\mathbf{Q}_2}) \pm \frac{1}{2}\sqrt{(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{Q}_2})^2 + 4(mU)^2}. \quad (23)$$

A gap opens where the bands $\varepsilon_{\mathbf{k}}$ and $\varepsilon_{\mathbf{k}+\mathbf{Q}_2}$ cross, e.g., at the X point of the original Brillouin zone (Fig. 12). At half filling and for $mU = 0$ the Fermi level crosses the bands at the X point too; thus the system is insulator for any finite value of mU . In the small t/U limit we can assume that $m = 1/2$ and expand the eigenvalues in powers of $\varepsilon_{\mathbf{k}}/U$. For the occupied states we find

$$\varepsilon_{\mathbf{k}-} - \mu \sim -\frac{1}{2}U - \frac{\varepsilon_{\mathbf{k}}^2}{U} = -\frac{1}{2}U - \frac{4t^2}{U} \left(\frac{\varepsilon_{\mathbf{k}}}{2t}\right)^2$$

The ground-state total energy for the antiferromagnetic supercell is then $2E_{\text{AF}}$ with

$$E_{\text{AF}} = -\frac{1}{2}U - \frac{4t^2}{U} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left(\frac{\varepsilon_{\mathbf{k}}}{2t}\right)^2 \sim -\frac{1}{2}U - \frac{4t^2}{U}$$

so that the energy difference per couple of spins between ferro- and antiferro-magnetic state is

$$\Delta E^{\text{HF}} = E_{\uparrow\uparrow}^{\text{HF}} - E_{\uparrow\downarrow}^{\text{HF}} = \frac{2}{n_{\langle ii' \rangle}} [E_{\text{F}} - E_{\text{AF}}] \sim \frac{1}{2} \frac{4t^2}{U} \sim \frac{1}{2} \Gamma, \quad (24)$$

which is similar to the result obtained from the Hubbard model in many-body second order perturbation theory, Eq. (12). Despite of the similarity with the actual solution, one has to remember that the spectrum of the Hartree-Fock Hamiltonian has very little to do with the spectrum of the Heisenberg model, the model which describes the actual low-energy behavior of the Hubbard Hamiltonian. If we restrict ourselves to the antiferromagnetic solution, the first excited state is at an energy $\propto U$ rather than $\propto \Gamma$; thus we cannot use a single HF calculation to understand the magnetic excitation spectrum of a given system. It is more meaningful to use HF to compare the total energy of different states and determine in this way, within HF, the ground state. Even in this case, however, one has to keep in mind that HF suffers of *spin contamination*, i.e., singlet states and $S_z = 0$ triplet states mix [11]. The energy difference per bond $E_{\uparrow\uparrow}^{\text{HF}} - E_{\uparrow\downarrow}^{\text{HF}}$ in Eq. (24) only resembles the exact result, as one can grasp by comparing it with the actual energy difference between triplet and singlet state in the two-site Heisenberg model

$$\Delta E = E_{S=1} - E_{S=0} = \Gamma,$$

a factor two larger. The actual ratio $\Delta E/\Delta E^{\text{HF}}$ might depend on the details of the HF band structures. Thus, overall, Hartree-Fock is not the ideal approach to determine the onset of magnetic phase transitions. Other shortcomings of the Hartree-Fock approximation are in the description of the Mott metal-insulator transition. In Hartree-Fock the metal-insulator transition is intimately related to long-range magnetic order (Slater transition), but in strongly-correlated materials the metal-insulator transition can occur in the paramagnetic phase (Mott transition). It is associated with a divergence of the self-energy at low frequencies rather than with the formation of superstructures. This physics, captured by many-body methods such as the dynamical mean-field theory (DMFT) [6], is completely missed by the Hartree-Fock approximation.

3 The Kondo model

The Kondo impurity is a representative case of a system which exhibits both local moment and Pauli paramagnetic behavior, although in quite different temperature regimes [5]. The Kondo effect was first observed in diluted metallic alloys, metallic systems in which isolated d or f magnetic impurities are present, and it has been a riddle for decades. A Kondo impurity in a metallic host can be described by the Anderson model

$$H_A = \sum_{\sigma} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \sum_{\sigma} \varepsilon_f n_{f\sigma} + U n_{f\uparrow} n_{f\downarrow} + \sum_{\sigma} \sum_{\mathbf{k}} \left[V_{\mathbf{k}} c_{\mathbf{k}\sigma}^{\dagger} c_{f\sigma} + h.c. \right], \quad (25)$$

where ε_f is the impurity level, occupied by $n_f \sim 1$ electrons, $\varepsilon_{\mathbf{k}}$ is the dispersion of the metallic band and $V_{\mathbf{k}}$ the hybridization. If we assume that the system has particle-hole symmetry with respect to the Fermi level, then $\varepsilon_f = -U/2$. The Kondo regime is characterized by the parameters values $\varepsilon_f \ll \varepsilon_F$ and $\varepsilon_f + U \gg \varepsilon_F$, and by a weak hybridization, i.e., the hybridization function $\Delta(\varepsilon) = \pi \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} |V_{\mathbf{k}}|^2 \delta(\varepsilon_{\mathbf{k}} - \varepsilon)$ is such that $\Delta(\varepsilon_F) \ll |\varepsilon_F - \varepsilon_f|, |\varepsilon_F - \varepsilon_f - U|$. Through the Schrieffer-Wolff canonical transformation [12] one can map the Anderson model onto the Kondo model²

$$H_K = \sum_{\sigma} \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} + \Gamma \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) = H_0 + H_{\Gamma}, \quad (26)$$

where

$$\Gamma \sim -2|V_{k_F}|^2 \left[\frac{1}{\varepsilon_f} - \frac{1}{\varepsilon_f + U} \right] > 0$$

is the antiferromagnetic coupling arising from the hybridization, \mathbf{S}_f the spin of the impurity ($S_f = 1/2$), and $\mathbf{s}_c(\mathbf{0})$ is the spin-density of the conduction band at the impurity site. The solution of the problem defined by (25) or (26) is not at all trivial and requires many-body techniques such as the Wilson numerical renormalization group [14] or the Bethe Ansatz [15]. Here we only discuss some important exact results. First we define the *impurity susceptibility*, $\chi_{zz}^f(T)$, as the total susceptibility minus the contribution of the metallic band in the absence of the impurity [14–16]. One can show that at high temperatures $\chi_{zz}^f(T)$, has the following behavior

$$\chi_{zz}^f(T) \sim \frac{(g\mu_B)^2 S_f(S_f + 1)}{3k_B T} \left\{ 1 - \frac{1}{\ln(T/T_K)} \right\}.$$

This expression resembles the Curie susceptibility, apart from the $\ln(T/T_K)$ term. The scale T_K is the Kondo temperature, which, in first approximation, is given by

$$k_B T_K \sim D e^{-2/\rho(\varepsilon_F)\Gamma},$$

²The Schrieffer-Wolf transformation yields additionally a potential scattering interaction, a pair tunneling coupling and a shift of the energies $\varepsilon_{\mathbf{k}}$. These interactions are however not important for the discussion in this section and therefore we neglect them.

where $2D = W$ is the band-width of the host conduction band. Because of the $\ln(T/T_K)$ term, the susceptibility apparently diverges at $T \sim T_K$. In reality, however, around T_K there is a crossover to a new regime. For $T \ll T_K$

$$\chi_{zz}^f(T) \sim \frac{C_{1/2}}{\mathcal{W}T_K} \{1 - \alpha T^2 + \dots\},$$

where \mathcal{W} is a (universal) Wilson number. Thus the low-temperature system has a Fermi-liquid behavior with enhanced density of states, i.e., with heavy masses m^*/m ; furthermore $\chi_{zz}^f(0) = C_{1/2}/\mathcal{W}T_K$ is the Curie susceptibility (Eq. (15)) with the temperature *frozen* at $T = \mathcal{W}T_K$. At $T = 0$ the impurity magnetic moment is screened by the conduction electrons, which form a singlet state with the spin of the impurity. In other words, the effective magnetic moment formed by the impurity magnetic moment and its screening cloud,

$$\mu_{\text{eff}}^2(T) \equiv 3k_B T \chi_{zz}^f(T) \propto \langle S_z^f S_z^f \rangle + \langle S_z^f s_z^c \rangle,$$

vanishes for $T \ll T_K$. The Kondo temperature is typically 10-30 K or even smaller, hence the Fermi-liquid behavior is restricted to a very narrow energy and temperature region.

We can understand the existence of a Fermi-liquid regime by using a simple approach due to Anderson, the so-called *poor-man scaling* [17], and an argument due to Nozières. First we divide the Hilbert space into a high- and a low-energy sector. We define as *high-energy* states those with at least one electron or one hole at the top or bottom of the band; the corresponding constraint for the high-energy electronic level ε_q is

$$\begin{aligned} D' &< \varepsilon_q < D \\ -D &< \varepsilon_q < -D', \end{aligned}$$

where $D' = D - \delta D$. Next we introduce the operator P_H , which projects onto the high-energy states, and the operator $P_L = 1 - P_H$, which projects onto states with no electrons or holes in the high-energy region. Then we downfold the high-energy sector of the Hilbert space. To do this we rewrite the Kondo Hamiltonian as

$$\begin{aligned} H' &= P_L H P_L + \delta H_L = H_L + \delta H_L, \\ \delta H_L &= P_L H P_H (\omega - P_H H P_H)^{-1} P_H H P_L. \end{aligned}$$

Here H_L is the original Hamiltonian, however in the space in which the high-energy states have been eliminated; the term δH_L is a correction due to the interaction between low and (downfolded) high-energy states. Next we calculate δH_L using perturbation theory. The first non-zero contribution is of second order in Γ

$$\delta H_L^{(2)} \sim P_L H_\Gamma P_H (\omega - P_H H_0 P_H)^{-1} P_H H_\Gamma P_L.$$

There are two types of processes which contribute at this order, an electron and a hole process, depending if the downfolded states have (at least) one electron or one hole in the high-energy region. Let us consider the electron process. We set

$$P_H \sim \sum_{\sigma} \sum_{\mathbf{q}} c_{\mathbf{q}\sigma}^\dagger |FS\rangle \langle FS| c_{\mathbf{q}\sigma}, \quad P_L \sim \sum_{\sigma} \sum_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger |FS\rangle \langle FS| c_{\mathbf{k}\sigma},$$

where $|\varepsilon_k| < D'$ and $|FS\rangle = \prod_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}^\dagger |0\rangle$ is the Fermi sea, i.e., the many-body state corresponding to the metallic conduction band. Thus

$$\begin{aligned}\delta H_L^{(2)} &= -\frac{1}{2}I^2 \sum_q \frac{1}{\omega - \varepsilon_q} \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) + \dots \\ &\sim \frac{1}{4}\rho(\varepsilon_F)I^2 \frac{\delta D}{D} \mathbf{S}_f \cdot \mathbf{s}_c(\mathbf{0}) + \dots\end{aligned}$$

We find an analogous contribution from the hole process. The correction $\delta H_L^{(2)}$ modifies the parameter Γ of the Kondo Hamiltonian as follows

$$\begin{aligned}\Gamma &\rightarrow \Gamma' = \Gamma + \delta\Gamma, \\ \frac{\delta\Gamma}{\delta \ln D} &= \frac{1}{2}\rho(\varepsilon_F)\Gamma^2.\end{aligned}\tag{27}$$

The equation (27) has two fixed points, $\Gamma = 0$ (*weak coupling*) and $\Gamma \rightarrow \infty$ (*strong coupling*). By solving the scaling equations we find

$$\Gamma' = \frac{\Gamma}{1 + \frac{1}{2}\rho(\varepsilon_F)\Gamma \ln \frac{D'}{D}}.$$

If Γ is antiferromagnetic the renormalized coupling constant Γ' diverges for $D' = De^{-2/\Gamma\rho(\varepsilon_F)}$, an energy proportional to the Kondo energy $k_B T_K$. This divergence (scaling to strong coupling) indicates that at low energy the interaction between the spins dominates and therefore the system forms a singlet in which the impurity magnetic moment is screened. The existence of this strong coupling fixed point is confirmed by the numerical renormalization group of Wilson [14]. Nozières [19] has used this conclusion to show that the low-temperature behavior of the system must be of Fermi liquid type. His argument is the following. For infinite coupling Γ' the impurity traps a conduction electron to form a singlet state. For a finite but still very large Γ' any attempt of breaking the singlet will cost a very large energy. Virtual excitations (into the $n_f = 0$ or $n_f = 2$ states and finally the $n_f = 1$ triplet state) are however possible and they yield an effective indirect interaction between the remaining conduction electrons surrounding the impurity. This is similar to the phonon-mediated attractive interaction in metals. The indirect electron-electron coupling is weak and can be calculated in perturbation theory ($1/\Gamma$ expansion). Nozières has shown that, in first approximation, the effective interaction is between electrons of opposite spins lying next to the impurity, it is of order D^4/Γ^3 and repulsive; hence it gives rise to a Fermi-liquid behavior with enhanced susceptibility [19].

If $\Gamma = \Gamma_F < 0$ (ferromagnetic coupling, as for example the coupling arising from direct Coulomb exchange) the renormalized coupling constant Γ' goes to zero in the $D' \rightarrow 0$ limit (scaling to weak coupling). This means that the local spin becomes *asymptotically free* and yields a Curie-type susceptibility, which diverges for $T \rightarrow 0$. For small but finite coupling we can account for the ferromagnetic interaction perturbatively (expansion in orders of Γ_F). In f electron materials often both ferro and antiferromagnetic exchange couplings are present, the first, Γ_F , arising from the Coulomb exchange, the second, Γ , from the hybridization. There

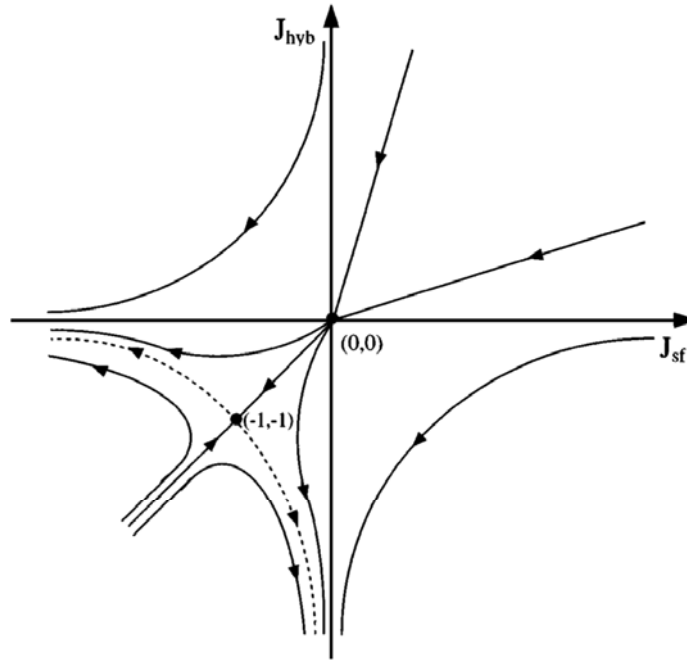


Fig. 13: Sketch of the scaling diagrams for the two-channel Kondo model. $\Gamma = -J_{\text{hyb}}$ and $\Gamma_F = -J_{\text{sf}}$. For $\Gamma > 0$ (antiferromagnetic) and $\Gamma_F < 0$ (ferromagnetic) the antiferromagnetic coupling scales to strong coupling and ferromagnetic one to weak coupling (right bottom quadrant). From Ref. [18].

are therefore two possibilities. If both exchange interactions couple the impurity with the same conduction channel, only the total coupling $\Gamma_F + \Gamma$ matters. Thus a $|\Gamma_F| > \Gamma$ suppresses the Kondo effect. If, however, ferromagnetic and antiferromagnetic exchange interaction couple the impurity to different conduction channels, a $|\Gamma_F| > \Gamma$ does not suppress the Kondo effect (Fig. 13), but merely reduces T_K . In the infinite $|\Gamma_F|$ limit the model describes an undercompensated Kondo effect [18].

4 Conclusion

In this lecture we introduced some of the fundamental aspects of magnetism in correlated systems. We have seen two distinct regimes, the itinerant and the local moment regime. In the first regime we can, in most cases, treat correlation effects in perturbation theory. In the world of real materials this is the limit in which the density-functional theory (DFT), in the local-density approximation or its simple extensions, works best. If the system is weakly correlated we can calculate the linear-response function in the random-phase approximation and understand fairly well magnetism within this approach.

The opposite regime is the strong-correlation regime, in which many-body effects play a key role. In this limit perturbation theory fails and we have in principle to work with many-body methods. If, however, we are interested only in magnetic phenomena, at integer filling a strong simplification comes from mapping the original many-body Hamiltonian into an effective spin

model. The exact solution of effective spins models requires in general numerical methods such as the Monte Carlo or quantum Monte Carlo approach, or, when the system is small enough, exact diagonalization or Lanczos. These techniques are discussed in the lectures of Werner Krauth, Stefan Wessel and Jürgen Schnack. The density-matrix renormalization group (DMRG), particularly efficient for one-dimensional systems, is instead presented in the lectures of Ulrich Schollwöck and Jens Eisert.

To work with material-specific spin models we need to calculate the magnetic exchange parameters. Typically this is done starting from total-energy DFT calculations for different spin configurations, e.g., in the LDA+ U approximation. The LDA+ U approach is based on the Hartree-Fock approximation, and therefore when we extract the parameters from LDA+ U calculations we have to keep in mind the shortcomings of the method. Furthermore if we want to extract the magnetic couplings from total energy calculations we have to make a guess on the form of the spin model. More flexible approaches, which allow us to account for actual correlations effects, are based on Green functions and the local-force theorem [20], as discussed in the lecture of Sasha Lichtenstein, or on canonical transformations [12, 21].

In strongly-correlated materials localized and itinerant moments physics can often be observed in the same system, although in different energy or temperature regimes. This is apparent in the case of the Kondo effect. For a Kondo impurity the susceptibility exhibits a Curie behavior at high temperature and a Fermi-liquid behavior at low temperature. In correlated transition-metal oxides Fermi liquid and local-spin magnetism can both play an important role but at different energy scales. Furthermore, in the absence of a large charge gap downfolding to spin models is not really justified. The modern method to bridge between localized and itinerant regime and deal with the actual complications of real systems is the dynamical mean-field theory (DMFT) [6]. Within this technique we solve directly generalized Hubbard-like models, however in the local self-energy approximation. DMFT is the first flexible approach that allows us to understand the paramagnetic Mott metal-insulator transition and thus also magnetism in correlated materials in a realistic setting. Modern DMFT codes are slowly but steadily becoming as complex and flexible as DFT codes, allowing us to deal with the full complexity of strongly-correlated materials. While this is a huge step forwards, we have to remember that state-of-the-art many-body techniques have been developed by solving simple models within certain approximations. We have to know very well these if we want to understand real materials and further advance the field. In DMFT we solve self-consistently an effective quantum-impurity model, a generalization of the Anderson model. Thus a lot can be learnt from the solution of the Anderson model in the context of the Kondo problem. Much can be understood alone with simple arguments, as Anderson or Nozières have shown us, reaching important conclusions on the Kondo problem with paper and pencil.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Formalism

The formulas in this Appendix are in atomic units: The numerical value of e , m and \hbar is 1, that of μ_B is 1/2, and energies are in Hartree.

A.1 Matsubara Green functions

A.1.1 Imaginary time and frequency Green functions

The imaginary time Matsubara Green function is defined as

$$G_{\alpha\beta}(\boldsymbol{\tau}) = -\langle \mathcal{T} c_{\alpha}(\tau_1) c_{\beta}^{\dagger}(\tau_2) \rangle = -\frac{1}{Z} \text{Tr} \left[e^{-\beta(H-\mu N)} \mathcal{T} c_{\alpha}(\tau_1) c_{\beta}^{\dagger}(\tau_2) \right],$$

where \mathcal{T} is the time-ordering operator, $\boldsymbol{\tau} = (\tau_1, \tau_2)$, $Z = \text{Tr} e^{-\beta(H-\mu N)}$ is the partition function, and the imaginary time operators $o(\tau) = c(\tau)$, $c^{\dagger}(\tau)$ are defined as $o(\tau) = e^{\tau(H-\mu N)} o e^{-\tau(H-\mu N)}$. The indices α and β are the flavors; they can be site and spin indices in the atomic limit, and \mathbf{k} and spin indices in the non-interacting electrons limit. Writing explicitly the action of the time-ordering operator we obtain

$$G_{\alpha\beta}(\boldsymbol{\tau}) = -\Theta(\tau_1 - \tau_2) \langle c_{\alpha}(\tau_1) c_{\beta}^{\dagger}(\tau_2) \rangle + \Theta(\tau_2 - \tau_1) \langle c_{\beta}^{\dagger}(\tau_2) c_{\alpha}(\tau_1) \rangle.$$

Using the invariance of the trace of the product of operators under cyclic permutations, one can show that the following properties hold

$$\begin{aligned} G_{\alpha\beta}(\boldsymbol{\tau}) &= G_{\alpha\beta}(\tau_1 - \tau_2), \\ G_{\alpha\beta}(\tau) &= -G_{\alpha\beta}(\tau + \beta) \quad \text{for } -\beta < \tau < 0. \end{aligned}$$

The Fourier transform on the Matsubara axis is

$$G_{\alpha\beta}(i\nu_n) = \frac{1}{2} \int_{-\beta}^{\beta} d\tau e^{i\nu_n \tau} G_{\alpha\beta}(\tau) = \int_0^{\beta} d\tau e^{i\nu_n \tau} G_{\alpha\beta}(\tau),$$

with $\nu_n = (2n + 1)\pi/\beta$. The inverse Fourier transform is given by

$$G_{\alpha\beta}(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{+\infty} e^{-i\nu_n \tau} G_{\alpha\beta}(i\nu_n).$$

The convergence of $G_{\alpha\beta}(\tau)$ is only guaranteed in the interval $-\beta < \tau < \beta$. A discussion of this can be found in the lecture of Robert Eder. Finally, if n_{α} is the number of electrons for flavor α , one can show that

$$G_{\alpha\alpha}(\tau \rightarrow 0^+) = -1 + n_{\alpha}, \quad G_{\alpha\alpha}(\tau \rightarrow \beta^-) = -n_{\alpha}. \quad (28)$$

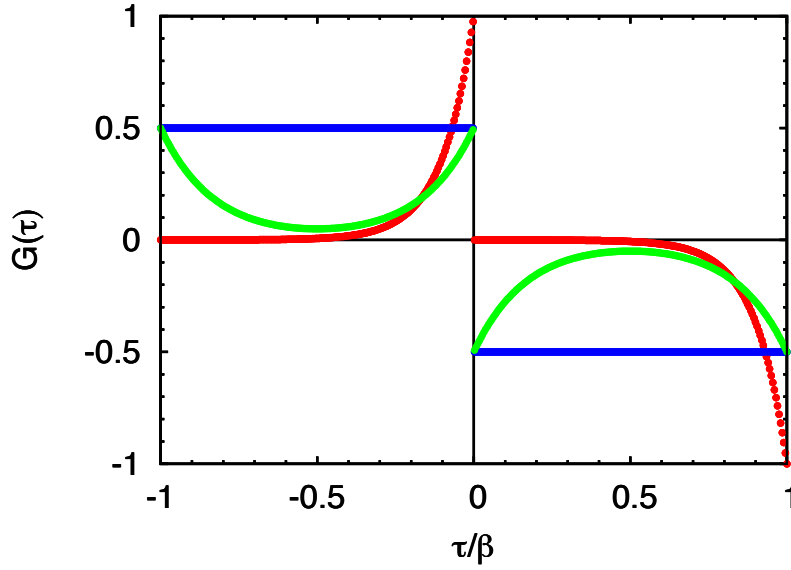


Fig. 14: The function $\mathcal{G}_{k\sigma}(\tau)$ defined in Eq. (30) for a state well below the Fermi level (red) and at the Fermi level (blue) and $\beta = 2 \text{ (eV)}^{-1}$. The green line shows the atomic $G(\tau)$ from Eq. (32) for $U = 6 \text{ eV}$ and $h = 0$.

A.1.2 Non-interacting limit

For a non-interacting system described by

$$H_0 = \sum_{\mathbf{k}} \sum_{\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} \quad (29)$$

we can show that

$$\begin{aligned} \mathcal{G}_{k\sigma}(\tau) &= -\langle \mathcal{T} [c_{k\sigma}(\tau) c_{k\sigma}^\dagger(0)] \rangle \\ &= -[\Theta(\tau) (1 - n_{\sigma}(\varepsilon_{\mathbf{k}})) - \Theta(-\tau) n_{\sigma}(\varepsilon_{\mathbf{k}})] e^{-(\varepsilon_{\mathbf{k}} - \mu)\tau}, \end{aligned} \quad (30)$$

where

$$n_{\sigma}(\varepsilon_{\mathbf{k}}) = \frac{1}{1 + e^{\beta(\varepsilon_{\mathbf{k}} - \mu)}}.$$

The Fourier transform at Matsubara frequencies is

$$\mathcal{G}_{k\sigma}(i\nu_n) = \frac{1}{i\nu_n - (\varepsilon_{\mathbf{k}} - \mu)}.$$

To obtain the analytic continuation of this Green function on the real axis we substitute

$$i\nu_n \rightarrow \omega + i0^+.$$

A.1.3 Matsubara sums

The non-interacting Green function $\mathcal{G}_{k\sigma}(z)$ has a pole at $z_p = \varepsilon_{\mathbf{k}} - \mu$; the Fermi function $n_{\sigma}(z)$ has instead poles for $z = i\nu_n$. Let us consider the integral

$$\frac{1}{2\pi i} \oint_C \mathcal{F}_{k\sigma}(z) n_{\sigma}(z) e^{z\tau} dz = 0,$$

where $0 < \tau < \beta$ and where the function $\mathcal{F}_{\mathbf{k}\sigma}(z)$ is analytic everywhere except at some poles $\{z_p\}$. The contour C is a circle in full complex plane centered at the origin and including the poles of the Fermi function (Matsubara frequencies) and the poles of $\mathcal{F}_{\mathbf{k}\sigma}(z)$. The integral is zero because the integrand vanishes exponentially for $|z| \rightarrow \infty$. Furthermore

$$\text{Res } [n_\sigma(i\nu_n)] = -\frac{1}{\beta}.$$

Using Cauchy's integral theorem we then have

$$\frac{1}{\beta} \sum_n e^{i\nu_n \tau} \mathcal{F}_{\mathbf{k}\sigma}(i\nu_n) = \sum_{z_p} \text{Res } [\mathcal{F}_{\mathbf{k}\sigma}(z_p)] n_\sigma(z_p) e^{z_p \tau}.$$

We can use this expression and (28) to show that

$$\begin{aligned} \frac{1}{\beta} \sum_n e^{-i\nu_n 0^-} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \mathcal{G}_{\mathbf{k}\sigma}(0^-) = n_\sigma(\varepsilon_{\mathbf{k}}), \\ \frac{1}{\beta} \sum_n e^{-i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \mathcal{G}_{\mathbf{k}\sigma}(0^+) = n_\sigma(\varepsilon_{\mathbf{k}}) - 1. \end{aligned}$$

In a similar way we can show that

$$\begin{aligned} \frac{1}{\beta} \sum_n e^{i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) &= \frac{dn_\sigma(\varepsilon_{\mathbf{k}})}{d\varepsilon_{\mathbf{k}}} = \beta n_\sigma(\varepsilon_{\mathbf{k}}) [-1 + n_\sigma(\varepsilon_{\mathbf{k}})], \\ \frac{1}{\beta} \sum_n e^{i\nu_n 0^+} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(i\nu_n + i\omega_m) &= \frac{n_{\mathbf{k}+\mathbf{q}} - n_{\mathbf{k}}}{i\omega_m + \varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}, \end{aligned}$$

where in the last relation $\omega_m = 2m\pi/\beta$ is a Bosonic Matsubara frequency.

A.1.4 Atomic limit

It is interesting to consider a half-filled idealized atom described by the Hamiltonian

$$H = \varepsilon_d \sum_\sigma n_\sigma + U \left(\frac{N^2}{4} - S_z^2 \right) + g\mu_B h S_z. \quad (31)$$

For $\tau > 0$ we can calculate explicitly the Green function obtaining

$$G_\sigma(\tau) = -\frac{1}{2} \frac{1}{1 + e^{\beta U/2} \cosh(\beta g\mu_B h/2)} \left[e^{\tau(U - g\mu_B h\sigma)/2} + e^{(\beta - \tau)(U + g\mu_B h\sigma)/2} \right]. \quad (32)$$

The Fourier transform of $G_\sigma(\tau)$ is

$$G_\sigma(i\nu_n) = \left[\frac{w_-}{i\nu_n + (U - g\mu_B h\sigma)/2} + \frac{w_+}{i\nu_n - (U + g\mu_B h\sigma)/2} \right],$$

where

$$w_\pm = \frac{1}{2} \frac{1 + e^{\beta U/2} e^{\pm \beta g\mu_B h\sigma/2}}{1 + e^{\beta U/2} \cosh(\beta g\mu_B h/2)}.$$

Since the Green function is written as the sum of functions with one pole, the analytic continuation is simple, as in the non-interacting case. We replace $i\nu_n$ with $\omega + i0^+$.

A.1.5 Lehmann representation

Using the Lehmann representation we can rewrite

$$G_{\mathbf{k}\sigma}(i\nu_n) = \int A_{\mathbf{k}\sigma}(\varepsilon) \frac{1}{i\nu_n - \varepsilon} d\varepsilon,$$

where $A_{\mathbf{k}\sigma}(\varepsilon) = -\frac{1}{\pi} \text{Im} [G_{\mathbf{k}\sigma}(\varepsilon)]$ is the spectral function. The spectral function is related to the density of states as follows

$$\rho_\sigma(\varepsilon) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} A_{\mathbf{k}\sigma}(\varepsilon).$$

A.2 Linear response theory

A.2.1 Theory

The response of a system described by the Hamiltonian H to a small magnetic field $\mathbf{h}(\mathbf{r}, t)$ is given by the linear correction to the Hamiltonian, i.e.,

$$\sum_{\nu} \delta H_{\nu}(\mathbf{r}; t) = - \sum_{\nu} M_{\nu}(\mathbf{r}; t) h_{\nu}(\mathbf{r}; t), \quad (33)$$

where $M(\mathbf{r}; t)$ is the magnetization operator in the Heisenberg representation

$$M_{\nu}(\mathbf{r}; t) = e^{iHt} M_{\nu}(\mathbf{r}) e^{-iHt}$$

and $\nu = x, y, z$. To linear order in the perturbation, and assuming that the perturbation is turned on adiabatically at $t_0 = -\infty$

$$\langle M_{\nu}(\mathbf{r}; t) \rangle = \langle M_{\nu}(\mathbf{r}) \rangle_0 - i \sum_{\nu'} \int d\mathbf{r}' \int_{-\infty}^t dt' \langle [M_{\nu}(\mathbf{r}; t), \delta H_{\nu'}(\mathbf{r}'; t')] \rangle_0,$$

where $\langle M_{\nu}(\mathbf{r}) \rangle_0$ is the (equilibrium) thermal average in the absence of the perturbation. By replacing $\sum_{\nu'} \delta H_{\nu'}(\mathbf{r}'; t')$ with the expression (33) we obtain

$$\delta \langle M_{\nu}(\mathbf{r}; t) \rangle = \langle M_{\nu}(\mathbf{r}; t) \rangle - \langle M_{\nu}(\mathbf{r}) \rangle_0 = i \sum_{\nu'} \int d\mathbf{r}' \int_{-\infty}^t dt' \langle [M_{\nu}(\mathbf{r}; t), M_{\nu'}(\mathbf{r}'; t')] \rangle_0 h_{\nu'}(\mathbf{r}'; t').$$

The function

$$\chi_{\nu\nu'}(\mathbf{r}, \mathbf{r}'; t, t') = i \langle [M_{\nu}(\mathbf{r}; t), M_{\nu'}(\mathbf{r}'; t')] \rangle_0 \Theta(t - t') \quad (34)$$

is the so-called retarded response function. It is often convenient to work with the operators $\Delta M_{\nu}(\mathbf{r}; t) = M_{\nu}(\mathbf{r}; t) - \langle M_{\nu}(\mathbf{r}) \rangle_0$ which measure the deviation with respect to the average in the absence of perturbation; since numbers always commute, we can replace $M_{\nu}(\mathbf{r}; t)$ with $\Delta M_{\nu}(\mathbf{r}; t)$ in the expression (34).

If the Hamiltonian H has time translational invariance symmetry the retarded response function depends only on time differences $t - t'$. For the Fourier transform we have

$$\delta\langle M_\nu(\mathbf{r}; \omega) \rangle = \sum_{\nu'} \int d\mathbf{r}' \chi_{\nu\nu'}(\mathbf{r}, \mathbf{r}'; \omega) h_{\nu'}(\mathbf{r}'; \omega).$$

For a lattice with lattice translational invariance, if we Fourier transform to reciprocal space and integrate over the unit cell

$$\delta\langle M_\nu(\mathbf{q}; \omega) \rangle = \sum_{\nu'} \int d\mathbf{r} \int d\mathbf{r}' e^{i\mathbf{q}\cdot\mathbf{r}} \chi_{\nu\nu'}(\mathbf{q}, \mathbf{r}, \mathbf{r}'; \omega) h_{\nu'}(\mathbf{q}, \mathbf{r}'; \omega).$$

Finally, if the perturbation depends on \mathbf{r}' only through a phase we obtain

$$\delta\langle M_\nu(\mathbf{q}; \omega) \rangle = \sum_{\nu'} \int d\mathbf{r} \int d\mathbf{r}' e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \chi_{\nu\nu'}(\mathbf{q}, \mathbf{r}, \mathbf{r}'; \omega) h_{\nu'}(\mathbf{q}; \omega) = \sum_{\nu'} \chi_{\nu\nu'}(\mathbf{q}; \omega) h_{\nu'}(\mathbf{q}; \omega).$$

In the $\omega = 0$ and $\mathbf{q} \rightarrow \mathbf{0}$ limit we have

$$\chi_{\nu\nu'}(\mathbf{0}; 0) = \lim_{h_{\nu'} \rightarrow 0} \frac{\partial M_\nu}{\partial h_{\nu'}},$$

where $h_{\nu'} = h_{\nu'}(\mathbf{0}; 0)$.

A.2.2 Kramers-Kronig relations and thermodynamic sum rule

Important properties of the spin susceptibility are the Kramers-Kronig relations

$$\begin{aligned} \text{Re}[\chi(\mathbf{q}; \omega)] - \text{Re}[\chi(\mathbf{q}; \infty)] &= \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Im}[\chi(\mathbf{q}; \omega')]}{\omega' - \omega} d\omega', \\ \text{Im}[\chi(\mathbf{q}; \omega)] &= -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Re}[\chi(\mathbf{q}; \omega')] - \text{Re}[\chi(\mathbf{q}; \infty)]}{\omega' - \omega} d\omega', \end{aligned}$$

where \mathcal{P} is the Cauchy principal value, and Re and Im indicate the real and imaginary part.

The first Kramers-Kronig relation yields the sum rule

$$\text{Re}[\chi(\mathbf{q}; \omega = 0)] - \text{Re}[\chi(\mathbf{q}; \infty)] = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{+\infty} \frac{\text{Im}[\chi(\mathbf{q}; \omega')]}{\omega'} d\omega'. \quad (35)$$

In the $\mathbf{q} = \mathbf{0}$ limit, Eq. (35) is known as *thermodynamic sum rule*.

A.2.3 Fluctuation-dissipation theorem and static susceptibility

We define the spin correlation function

$$\begin{aligned} \mathcal{S}_{\nu\nu'}(\mathbf{q}; t) &= \langle \Delta S_\nu(\mathbf{q}; t) \Delta S_{\nu'}(-\mathbf{q}) \rangle_0 \\ &= \langle S_\nu(\mathbf{q}; t) S_{\nu'}(-\mathbf{q}) \rangle_0 - \langle S_\nu(\mathbf{q}) \rangle_0 \langle S_{\nu'}(-\mathbf{q}) \rangle_0 \end{aligned}$$

where $\Delta S_\nu(\mathbf{q}; t) = S_\nu(\mathbf{q}; t) - \langle S_\nu(\mathbf{q}) \rangle_0$ and $\langle S_\nu(\mathbf{q}) \rangle_0 = \langle S_\nu(\mathbf{q}; 0) \rangle_0$. The fluctuation-dissipation theorem relates $\mathcal{S}_{\nu\nu'}(\mathbf{q}; t)$ with the magnetic susceptibility. First, one can show that the following relation holds

$$\text{Im}[\chi_{\nu\nu'}(\mathbf{q}; t)] = (g\mu_B)^2 \frac{1}{2} [\mathcal{S}_{\nu\nu'}(\mathbf{q}; t) - \mathcal{S}_{\nu'\nu}(\mathbf{q}; -t)]. \quad (36)$$

The correlation function has the property

$$\mathcal{S}_{\nu\nu'}(\mathbf{q}; \omega) = e^{\beta\omega} \mathcal{S}_{\nu'\nu}(\mathbf{q}; -\omega).$$

Thus, in ω space Eq. (36) is replaced by

$$\text{Im}[\chi_{\nu\nu'}(\mathbf{q}; \omega)] = \frac{1}{2(1 + n_B)} (g\mu_B)^2 \mathcal{S}_{\nu\nu'}(\mathbf{q}; \omega), \quad n_B(\omega) = \frac{1}{e^{\beta\omega} - 1}.$$

Assuming $k_B T$ large and using Eq. (35) one can then show that

$$\text{Re}[\chi_{\nu\nu'}(\mathbf{q}; \omega = 0)] - \text{Re}[\chi_{\nu\nu'}(\mathbf{q}; \infty)] \sim \frac{(g\mu_B)^2}{k_B T} \mathcal{S}_{\nu\nu'}(\mathbf{q}; t = 0).$$

A.2.4 Imaginary time and frequency response function

We define the susceptibility in imaginary time as

$$\begin{aligned} \chi_{\nu\nu'}(\mathbf{q}; \tau, \tau') &= \langle \mathcal{T} \Delta M_\nu(\mathbf{q}; \tau) \Delta M_{\nu'}(-\mathbf{q}; \tau') \rangle_0 \\ &= \langle \mathcal{T} M_\nu(\mathbf{q}; \tau) M_{\nu'}(-\mathbf{q}; \tau') \rangle_0 - \langle M_\nu(\mathbf{q}) \rangle_0 \langle M_{\nu'}(-\mathbf{q}) \rangle_0, \end{aligned}$$

where $\Delta M_\nu(\mathbf{q}; \tau) = M_\nu(\mathbf{q}; \tau) - \langle M_\nu(\mathbf{q}; \tau) \rangle_0 = M_\nu(\mathbf{q}; \tau) - \langle M_\nu(\mathbf{q}) \rangle_0$. As in the case of the Green function, by using the invariance properties of the trace one can show that

$$\chi_{\nu\nu'}(\mathbf{q}; \tau, \tau') = \chi_{\nu\nu'}(\mathbf{q}; \tau - \tau').$$

The response function in imaginary time is related to the response function at the Bosonic Matsubara frequency $i\omega_n$ through the Fourier transforms

$$\chi_{\nu\nu'}(\mathbf{q}; \tau) = \frac{1}{\beta} \sum_n e^{-i\omega_n \tau} \chi_{\nu\nu'}(\mathbf{q}; i\omega_n),$$

$$\chi_{\nu\nu'}(\mathbf{q}; i\omega_n) = \int d\tau e^{i\omega_n \tau} \chi_{\nu\nu'}(\mathbf{q}; \tau).$$

In the rest of the Appendix we replace for simplicity the notation $\langle \rangle_0$ with $\langle \rangle$.

A.3 Magnetic susceptibility

A.3.1 Spin and magnetization operators

The spin operators S_ν are defined as

$$S_\nu = \frac{1}{2} \sum_{\sigma\sigma'} c_\sigma^\dagger \sigma_\nu c_{\sigma'},$$

where $\nu = x, y, z$ and σ_ν are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The magnetization operators M_ν are defined as $M_\nu = -g\mu_B S_\nu$.

A.3.2 Matsubara magnetic susceptibility

The magnetic susceptibility for a single-band system can be expressed as

$$\chi_{zz}(\mathbf{q}; \boldsymbol{\tau}) = (g\mu_B)^2 \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}), \quad (37)$$

where $\sigma = 1$ or -1 depending if the spin is up or down, $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$ and

$$\begin{aligned} \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}) &= \frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \chi_{\mathbf{k}}^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}), \\ \chi_{\mathbf{k}}^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}) &= \langle \mathcal{T} c_{\mathbf{k}\sigma}(\tau_1) c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger(\tau_2) c_{\mathbf{k}+\mathbf{q}\sigma'}(\tau_3) c_{\mathbf{k}\sigma'}^\dagger(\tau_4) \rangle \\ &\quad - \langle \mathcal{T} c_{\mathbf{k}\sigma}(\tau_1) c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger(\tau_2) \rangle \langle \mathcal{T} c_{\mathbf{k}+\mathbf{q}\sigma'}(\tau_3) c_{\mathbf{k}\sigma'}^\dagger(\tau_4) \rangle. \end{aligned}$$

In Fourier space

$$\chi_{zz}(\mathbf{q}; i\omega_m) = \frac{1}{4} \sum_{\sigma\sigma'} \sigma\sigma' \frac{1}{\beta^2} \sum_{nn'} \chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m),$$

where $\omega_m = 2m\pi/\beta$ is a Bosonic Matsubara frequency and

$$\chi_{n,n'}^{\mathbf{q}\sigma\sigma'}(i\omega_m) = \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\nu}) = \frac{1}{16} \iiint d\boldsymbol{\tau} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\tau}} \chi^{\mathbf{q}\sigma\sigma'}(\boldsymbol{\tau}).$$

The integral for each $\boldsymbol{\tau}$ component is from $-\beta$ to β and $\boldsymbol{\nu} = (\nu_n, -\nu_n - \omega_m, \nu_{n'} + \omega_m, -\nu_{n'})$.

A.3.3 Generalized Matsubara two-particle Green function

We define the generalized two-particle Green function

$$\chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\tau}) = \langle \mathcal{T} c_\alpha(\tau_1) c_\beta^\dagger(\tau_2) c_\gamma(\tau_3) c_\delta^\dagger(\tau_4) \rangle - \langle \mathcal{T} c_\alpha(\tau_1) c_\beta^\dagger(\tau_2) \rangle \langle \mathcal{T} c_\gamma(\tau_3) c_\delta^\dagger(\tau_4) \rangle. \quad (38)$$

The Fourier transform of (38) is

$$\chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\nu}) = \chi_{n,n'}^{\alpha\beta\gamma\delta}(i\omega_m) = \frac{1}{16} \iiint d\boldsymbol{\tau} e^{i\boldsymbol{\nu}\cdot\boldsymbol{\tau}} \chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\tau}).$$

From the symmetry properties of the trace we find that $\nu_4 = -\nu_1 - \nu_2 - \nu_3$. If we redefine $\nu_1 = \nu_n$, $\nu_2 = -\nu_n - \omega_m$, $\nu_3 = \nu_{n'} + \omega_m$, and $\nu_4 = -\nu_{n'}$ we obtain

$$\begin{aligned} \boldsymbol{\nu} &= (\nu_n, -\nu_n - \omega_m, \nu_{n'} + \omega_m, -\nu_{n'}), \\ \chi_{n,n'}^{\alpha\beta\gamma\delta}(i\omega_m) &= \frac{1}{16} \iiint d\boldsymbol{\tau} e^{i[-\omega_m\tau_{23} + \nu_n\tau_{12} + \nu_{n'}\tau_{34}]} \chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\tau}), \end{aligned} \quad (39)$$

where $\tau_{ij} = \tau_i - \tau_j$. The complex conjugate is given by

$$\left[\chi_{n,n'}^{\alpha\beta\gamma\delta}(i\omega_m) \right]^* = \chi_{-n-1, -n'-1}^{\alpha\beta\gamma\delta}(-i\omega_m),$$

where $\nu_{-n-1} = -\nu_n$, and $\nu_{-n'-1} = -\nu_{n'}$.

A.3.4 Symmetry properties

Let us now analyze the symmetry properties of (39). By using the fact that the response function (38) is real in τ space and by exchanging the indices 1 and 4, 2 and 3 in the integrand, we find

$$\chi_{n,n'}^{\alpha\beta\gamma\delta}(i\omega_m) = \chi_{n',n}^{\delta\gamma\beta\alpha}(i\omega_m),$$

and hence, if $\alpha = \delta$, $\beta = \gamma$, $\nu_n = \nu_{n'}$ is a reflection axis for the absolute value of (39)

$$\left| \chi_{n,n'}^{\alpha\beta\gamma\delta}(i\omega_m) \right| = \left| \chi_{n',n}^{\delta\gamma\beta\alpha}(i\omega_m) \right|$$

An additional reflection axis can be found by first shifting the frequency $\nu_n = \nu_l - \omega_m$

$$\chi_{l,n'}^{\alpha\beta\gamma\delta}(i\omega_m) = \frac{1}{16} \iiint d\boldsymbol{\tau} e^{i(-\omega_m\tau_{13} + \nu_l\tau_{12} + \nu_{n'}\tau_{34})} \chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\tau})$$

and then exchanging in the integrand the indices 12 with 34 and vice versa. Hence

$$\chi_{l,n'}^{\alpha\beta\gamma\delta}(i\omega_m) = \chi_{n',l}^{\gamma\delta\alpha\beta}(-i\omega_m)$$

so that, if $\alpha = \gamma$ and $\beta = \delta$, $\nu_{n+m} = -\nu_{n'}$ is a mirror line

$$\left| \chi_{n+m,n'}^{\alpha\beta\gamma\delta}(i\omega_m) \right| = \left| \chi_{-n'-1, -n-m-1}^{\gamma\delta\alpha\beta}(i\omega_m) \right|.$$

A.3.5 Non interacting limit

For a non-interacting system Wick's theorem yields

$$\begin{aligned} \chi_{\gamma\delta}^{\alpha\beta}(\boldsymbol{\tau}) &= -\langle \mathcal{T} c_\alpha(\tau_1) c_\delta^\dagger(\tau_4) \rangle \langle \mathcal{T} c_\gamma(\tau_3) c_\beta^\dagger(\tau_2) \rangle \\ &= -G_{\alpha\delta}(\tau_{14}) G_{\gamma\beta}(-\tau_{23}). \end{aligned}$$

We take as example the one band model (29) and set $\alpha = \mathbf{k}\sigma$, $\beta = \mathbf{k} + \mathbf{q}\sigma$, $\gamma = \mathbf{k} + \mathbf{q}\sigma'$, and $\delta = \mathbf{k}\sigma'$. Then, in the paramagnetic case, the magnetic susceptibility is given by

$$\chi_{zz}(\mathbf{q}; \boldsymbol{\tau}) = -(g\mu_B)^2 \frac{1}{4} \frac{1}{\beta} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \sum_{\sigma} \mathcal{G}_{\mathbf{k}\sigma}(\tau_{14}) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(\tau_{32}).$$

Its Fourier transform is

$$\chi_{zz}(\mathbf{q}; i\omega_m) = (g\mu_B)^2 \frac{1}{4} \frac{1}{\beta^2} \sum_{nn'} \sum_{\sigma} \chi_{n,n'}^{\mathbf{q}\sigma\sigma}(i\omega_m),$$

where

$$\sum_{\sigma} \chi_{n,n'}^{\mathbf{q}\sigma\sigma}(i\omega_m) = -\beta \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \sum_{\sigma} \mathcal{G}_{\mathbf{k}\sigma}(i\nu_n) \mathcal{G}_{\mathbf{k}+\mathbf{q}\sigma}(i\nu_n + i\omega_m) \delta_{n,n'}.$$

Thus the static susceptibility is

$$\chi_{zz}(\mathbf{q}; 0) = -(g\mu_B)^2 \frac{1}{4} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \sum_{\sigma} \frac{n_{\sigma}(\varepsilon_{\mathbf{k}+\mathbf{q}}) - n_{\sigma}(\varepsilon_{\mathbf{k}})}{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}.$$

Finally, in the $\mathbf{q} \rightarrow 0$ and $T \rightarrow 0$ limit we find

$$\begin{aligned} \chi_{zz}(\mathbf{0}; 0) &= \frac{1}{4} (g\mu_B)^2 \rho(\varepsilon_F) \\ \rho(\varepsilon_F) &= -\sum_{\sigma} \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left. \frac{dn_{\sigma}(\varepsilon_{\mathbf{k}})}{d\varepsilon_{\mathbf{k}}} \right|_{T=0} \end{aligned}$$

A.3.6 Atomic limit

We calculate the local atomic susceptibility for the system described by the Hamiltonian (31) starting from the general expression (38). In the sector $\tau_i > \tau_{i+1}$ we have

$$\chi_{\sigma'\sigma'}^{\sigma\sigma}(\boldsymbol{\tau}) = \frac{1}{2(1 + e^{\beta U/2})} \left(e^{\tau_{12}U/2 + \tau_{34}U/2} + \delta_{\sigma\sigma'} e^{(\beta - \tau_{12})U/2 - \tau_{34}U/2} \right).$$

The magnetic susceptibility for $\tau_i > \tau_{i+1}$ is then given by

$$\chi_{zz}(\boldsymbol{\tau}) = (g\mu_B)^2 \frac{1}{4} \frac{1}{\beta} \sum_{\sigma\sigma'} \sigma\sigma' \chi_{\sigma'\sigma'}^{\sigma\sigma}(\boldsymbol{\tau}) = \frac{(g\mu_B)^2}{4\beta} \frac{1}{(1 + e^{\beta U/2})} e^{(\beta - \tau_{12} - \tau_{34})U/2},$$

which depends only on $\tau_{12} + \tau_{34}$. If we perform the Fourier transform we recover formula for the uniform static susceptibility

$$\chi_{zz}(\mathbf{0}; 0) = (g\mu_B)^2 \frac{1}{4k_B T} \frac{e^{\beta U/2}}{1 + e^{\beta U/2}} = (g\mu_B)^2 \frac{1}{4} \frac{1}{\beta^2} \sum_{nn'} \sum_{\sigma\sigma'} \sigma\sigma' \chi_{n,n'}^{\mathbf{0}\sigma\sigma'}(0),$$

where

$$\sum_{\sigma\sigma'} \sigma\sigma' \chi_{n,n'}^{\mathbf{0}\sigma\sigma'}(i\omega_m) = \beta \frac{1}{i\nu_n - U/2} \frac{1}{i\nu_{n'} - U/2} \frac{1 + e^{\beta U/2}}{e^{\beta U/2}} \delta_{\omega_m, 0}.$$

A.3.7 Alternative formulation

The spin susceptibility can also be obtained from $\chi_{zz}^{i,i'}(\boldsymbol{\tau})$ with $\boldsymbol{\tau} = (\tau_1, \tau_1, \tau_3, \tau_3)$. We have

$$\chi_{zz}^{i,i'}(\boldsymbol{\tau}) = \chi_{zz}^{i,i'}(\tau_{13}) = \langle \mathcal{T} M_z^i(\tau_1) M_z^{i'}(\tau_3) \rangle - \langle M_z^i \rangle \langle M_z^{i'} \rangle,$$

where $M_z^i = -g\mu_B S_z^i$ is the magnetization at site i . Its Fourier transform is

$$\chi_{zz}^{i,i'}(i\omega_m) = \int d\tau_{13} e^{i\omega_m \tau_{13}} \chi_{zz}^{i,i'}(\tau_{13}).$$

By Fourier transforming to the reciprocal space we find

$$\begin{aligned} \chi_{zz}(\mathbf{q}; \tau_{13}) &= \langle \mathcal{T} M_z(\mathbf{q}; \tau_1) M_z(-\mathbf{q}; \tau_3) \rangle - \langle M_z(\mathbf{q}) \rangle \langle M_z(-\mathbf{q}) \rangle, \\ \chi_{zz}(\mathbf{q}; i\omega_m) &= \int d\tau e^{i\omega_m \tau_{13}} \chi_{zz}(\mathbf{q}; \tau_{13}). \end{aligned}$$

References

- [1] H. Bruus and K. Flensberg: *Many-Body Quantum Theory in Condensed Matter Physics* (Oxford University Press, 2004)
- [2] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)
- [3] D.C. Matthis: *The Theory of Magnetism Made Simple* (World Scientific, Singapore, 2006)
- [4] K. Yosida: *Theory of Magnetism* (Springer, Heidelberg, 1998)
- [5] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge University Press, 1993)
- [6] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt:
The LDA+DMFT approach to strongly correlated materials,
Reihe Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich, 2011)
<http://www.cond-mat.de/events/correl11>
- [7] E. Pavarini, E. Koch, F. Anders, M. Jarrell:
Correlated electrons: from models to materials,
Reihe Modeling and Simulation, Vol. 2 (Forschungszentrum Jülich, 2012)
<http://www.cond-mat.de/events/correl12>
- [8] See R. Eder, *Multiplets in Transition Metal Ions* in Ref. [7]
- [9] See E. Pavarini, *Crystal-field Theory, Tight-binding Method and Jahn-Teller Effect*,
in Ref. [7]
- [10] See E. Pavarini, *The LDA+DMFT Approach*, in Ref. [6]
- [11] See E. Koch, *Exchange Mechanisms*, in Ref. [7]
- [12] J.R. Schrieffer and P.A. Wolff, Phys. Rev. **149**, 491 (1966);
A.H. MacDonald, S.M. Girvin and D. Yoshioka, Phys. Rev. B **37**, 9753 (1988)
- [13] E. Pavarini, I. Dasgupta, T.Saha-Dasgupta, O. Jepsen and O.K. Andersen,
Phys. Rev. Lett. **87**, 047003 (2001)
- [14] K. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
- [15] N. Andrei, K. Furuya, and J.H. Lowenstein, Rev. Mod. Phys. **55**, 331 (1983);
A.M. Tsvelik and P.B. Wiegmann, Adv. Phys. **32**, 453 (1983)

- [16] J.E. Gubernatis, J.E. Hirsch, and D.J. Scalapino,
Phys. Rev. B **16**, 8478 (1987)
- [17] P.W. Anderson, J. Phys. C: Solid State Phys. **3**, 2436 (1970)
- [18] E. Pavarini and L.C. Andreani, Phys. Rev. Lett. **77**, 2762 (1996)
- [19] P. Nozières, J. Low. Temp. Phys. **17**, 31 (1974)
- [20] A.I. Lichtenstein, M.I. Katsnelson, and V.A. Gubanov, J. Phys. F **14**, L125;
Solid State Commun. **54**, 327 (1985);
A.I. Lichtenstein, M.I. Katsnelson, V.P. Antropov, and V.A. Gubanov,
J. Magn. Magn. Mater. **67**, 65 (1987);
M.I. Katsnelson and A.I. Lichtenstein, Phys. Rev. B **61**, 8906 (2000)
- [21] E. Pavarini, E. Koch, and A.I. Lichtenstein,
Phys. Rev. Lett. **101**, 266405 (2008);
A. Chiesa, S. Carretta, P. Santini, G. Amoretti and E. Pavarini,
Phys. Rev. Lett. **110**, 157204 (2013)

4 The Variational Cluster Approximation

Robert Eder

Institut für Festkörperphysik

Karlsruhe Institute of Technology

Contents

1	Introduction	2
2	Notation and brief review of field theory	3
3	Proof of the theorem by Luttinger and Ward	7
3.1	Statement of the theorem	7
3.2	The case $\lambda = 0$	8
3.3	Calculation of $\partial\Omega/\partial\lambda$	9
3.4	Definition and properties of the Luttinger-Ward functional	10
3.5	Calculation of $\partial\tilde{\Omega}/\partial\lambda$	15
4	The variational cluster approximation	17
5	Applications of the VCA	21
5.1	Metal-insulator transition in a dimer	21
5.2	Discussion of spontaneous symmetry breaking	25
5.3	Photoemission spectra of NiO, CoO and MnO	26
6	Summary	28
7	Appendix: A theorem on determinants	29

1 Introduction

A quantity of central importance for the description of correlated electron systems is the electronic self-energy $\Sigma(\mathbf{k}, \omega)$. It may be viewed as a momentum- and energy-dependent correction to the energy of an electron that describes the effects of its interaction with the other electrons. Here the word ‘correction’ is by no means supposed to imply that $\Sigma(\mathbf{k}, \omega)$ is small. Quite the contrary, for example, in a Mott-insulator $\Sigma(\mathbf{k}, \omega)$ contains a term of the form $U/\hbar(\omega - \omega_0)$, with U the intra-atomic Coulomb repulsion, and this term is both large and strongly dependent on the energy ω . In fact, the very reason why density functional calculations do not reproduce the single-particle excitation spectrum – or ‘band structure’ – of Mott-insulators is that they implicitly assume an ω -independent self-energy and thus miss this ‘correction’ of order U .

It would therefore seem desirable to have a theoretical principle that allows us to actually compute the self-energy of a correlated electron system, and in fact it can be shown that $\Sigma(\mathbf{k}, \omega)$ obeys a stationarity condition that can be used for that purpose. More precisely, Luttinger and Ward have shown in a seminal paper [1] that the grand canonical potential Ω of an interacting fermion system can be expressed as a functional of its self-energy, $\Omega = \Omega[\Sigma]$, and that this functional is stationary with respect to variations of Σ :

$$\frac{\delta \Omega}{\delta \Sigma(\mathbf{k}, \omega)} = 0 .$$

Unfortunately a straightforward application of the stationarity property – e.g. by introducing ‘trial self-energies’ that depend on a number of variational parameters – is not possible because $\Omega[\Sigma]$ involves the so-called Luttinger-Ward functional $F[\Sigma]$, which is defined as a sum over infinitely many Feynman diagrams and thus cannot be evaluated for a given trial self-energy. A possible approximation would be to truncate the Luttinger-Ward functional, thereby keeping only a selected class of Feynman diagrams, typically ‘bubbles’ or ‘ladders’. These are the famous conserving approximations of Baym and Kadanoff [2] and one example for such an approximation is the very successful GW-approximation proposed by Hedin [3]. On the other hand, the truncation of the Luttinger-Ward functional ultimately is a poorly controlled approximation that may be less suitable for strongly correlated electrons.

In 2003, however, an entirely new idea on how to apply the stationarity principle for Σ in strongly correlated electron systems was put forward by Potthoff, the so-called variational cluster approximation (VCA) [4–6]. The basic idea of the VCA is to generate trial self-energies Σ for an infinite lattice by exact diagonalization of finite clusters and, in the course of doing so, to evaluate the exact value of the Luttinger-Ward functional $F[\Sigma]$ numerically. Variation of Σ is performed by varying the single-particle terms of a cluster that serves as the ‘self-energy preparation-lab’. Put another way, the VCA seeks the best approximation to the self-energy of an infinite lattice amongst ‘cluster-representable’ ones, i.e. functions $\Sigma(\mathbf{k}, \omega)$ which can be generated as the exact self-energies of finite clusters. This is a new way for generating approximations in strongly correlated electron systems and in the following the variational principle itself, the basic idea of the VCA, and some selected applications will be presented.

2 Notation and brief review of field theory

First we define the notation and give a brief review of some concepts from field theory. While this will be rather cursory, introductions to the use of field theory in statistical physics can be found in many textbooks [7–9], in the present notes we try to be consistent with Fetter and Walecka (FW) [8].

We assume that the solid in question can be described as a periodic array of atomic orbitals centered on the nuclei of the atoms that form the basis of the lattice and we assume periodic boundary conditions. We choose the unit of length such that the unit cell has volume 1. All orbitals are taken as mutually orthogonal. The number of unit cells in the crystal is N and the number of atoms in the basis n_{Atom} . The orbitals can be labeled by a triple of indices (i, n, ν) where $i \in \{1, \dots, N\}$ denotes the unit cell, $n \in \{1, \dots, n_{Atom}\}$ the basis atom and $\nu \in \{s, p_x, p_y, p_z, d_{xy} \dots\}$ the type of orbital. The number of orbitals per unit cell is n_{orb} [10]. We introduce fermionic creation and annihilation operators, $c_{i,n,\nu,\sigma}^\dagger$ and $c_{i,n,\nu,\sigma}$, for electrons in these orbitals, where σ denotes the z -component of spin. It will often be convenient to contract (i, n, ν, σ) into a single ‘compound index’ α , so that the Hamiltonian – assumed to be time-independent – can be written as $H = H_0 + H_1$, with

$$H_0 = \sum_{\alpha,\beta} t_{\alpha,\beta} c_\alpha^\dagger c_\beta, \quad (1)$$

$$H_1 = \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} V_{\alpha,\beta,\delta,\gamma} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta. \quad (2)$$

Note the factor of $1/2$ and the ‘inverted’ order of indices on the interaction matrix element V in (2) which follows from the prescription of second quantization [7–9], see e.g. the lecture of E. Koch. The Fourier transform of the Fermion operators reads

$$c_{\mathbf{k},\beta}^\dagger = \frac{1}{\sqrt{N}} \sum_i e^{i\mathbf{k} \cdot (\mathbf{R}_i + \mathbf{r}_n)} c_{i,n,\nu,\sigma}^\dagger,$$

where we have introduced the orbital index $\beta = (n, \nu, \sigma)$. Since this second ‘compound index’ always comes together with either a momentum \mathbf{k} or a cell index i no misunderstanding is possible. The Hamiltonian now can be written as

$$H_0 = \sum_{\mathbf{k}} \sum_{\alpha,\beta} t_{\alpha,\beta}(\mathbf{k}) c_{\mathbf{k},\alpha}^\dagger c_{\mathbf{k},\beta}, \quad (3)$$

$$H_1 = \frac{1}{2N} \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} \sum_{\alpha,\beta,\gamma,\delta} V_{\alpha,\beta,\delta,\gamma}(\mathbf{k}, \mathbf{k}', \mathbf{q}) c_{\mathbf{k}+\mathbf{q},\alpha}^\dagger c_{\mathbf{k}'-\mathbf{q},\beta}^\dagger c_{\mathbf{k}',\gamma} c_{\mathbf{k},\delta}. \quad (4)$$

Equation (3) defines the $2n_{orb} \times 2n_{orb}$ matrix $t(\mathbf{k})$, whose eigenvalues $E_n(\mathbf{k})$ give the noninteracting band structure. This formulation allows H_0 to describe magnetic systems or include spin-orbit coupling. With the explicit prefactor of $1/N$ in (4) the matrix elements V in (4) are of order 1.

In all that follows we consider a grand canonical ensemble with inverse temperature $\beta = 1/k_B T$ and chemical potential μ . The thermal average of any operator \hat{O} is given by

$$\langle \hat{O} \rangle_{th} = \frac{1}{Z} \text{Tr} \left(e^{-\beta(H-\mu N)} \hat{O} \right) \quad (5)$$

with the grand partition function

$$Z = \text{Tr} \left(e^{-\beta(H-\mu N)} \right). \quad (6)$$

Introducing the imaginary-time Heisenberg operator

$$c_\alpha(\tau) = e^{\tau(H-\mu N)/\hbar} c_\alpha e^{-\tau(H-\mu N)/\hbar},$$

the *imaginary time Green's function* is defined as

$$G_{\alpha,\beta}(\tau) = -\Theta(\tau) \left\langle c_\alpha(\tau) c_\beta^\dagger \right\rangle_{th} + \Theta(-\tau) \left\langle c_\beta^\dagger c_\alpha(\tau) \right\rangle_{th} \quad (7)$$

$$\begin{aligned} &= \frac{1}{Z} \left(-\Theta(+\tau) \sum_{i,j} e^{-\beta(E_i-\mu N_i)} e^{\tau(E_i-E_j+\mu)/\hbar} \langle i|c_\alpha|j\rangle \langle j|c_\beta^\dagger|i\rangle \right. \\ &\quad \left. + \Theta(-\tau) \sum_{i,j} e^{-\beta(E_i-\mu N_i)} e^{\tau(E_j-E_i+\mu)/\hbar} \langle i|c_\beta^\dagger|j\rangle \langle j|c_\alpha|i\rangle \right), \end{aligned} \quad (8)$$

where $|i\rangle$ are the exact eigenstates of the system with energies E_i and particle number N_i and $\Theta(\tau)$ is the Heaviside step function. $\mathbf{G}(\tau)$ is a matrix with row dimension $2Nn_{orb}$, which can be made block-diagonal by introducing the spatial Fourier transform

$$G_{(n,\nu,\sigma),(n',\nu',\sigma')}(\mathbf{k}, \tau) = \frac{1}{N} \sum_{i,j} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j+\mathbf{r}_n-\mathbf{r}_{n'})} G_{(i,n,\nu,\sigma),(j,n',\nu',\sigma')}(\tau),$$

where $\mathbf{G}(\mathbf{k}, \tau)$ is a $2n_{orb} \times 2n_{orb}$ matrix.

From (8) it is easy to see that \mathbf{G} is well-defined only if $\tau \in [-\beta\hbar, \beta\hbar]$ when E_i are unbounded from above [11], and that for $\tau \in [-\beta\hbar, 0]$ one has $\mathbf{G}(\tau + \beta\hbar) = -\mathbf{G}(\tau)$. It follows that $\mathbf{G}(\tau)$ has the Fourier transform (see equation (25.10) in FW)

$$\begin{aligned} \mathbf{G}(\tau) &= \frac{1}{\beta\hbar} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu \tau} \mathbf{G}(i\omega_\nu), \\ \mathbf{G}(i\omega_\nu) &= \int_0^{\beta\hbar} d\tau e^{i\omega_\nu \tau} \mathbf{G}(\tau), \end{aligned}$$

The $\omega_\nu = (2\nu + 1)\pi/\beta\hbar$ are called the (Fermionic) Matsubara frequencies. From (8) we obtain

$$\begin{aligned} G_{\alpha,\beta}(i\omega_\nu) &= \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta(E_i-\mu N_i)} + e^{-\beta(E_j-\mu N_j)}}{i\omega_\nu + \frac{1}{\hbar}\mu - \frac{1}{\hbar}(E_j - E_i)} \langle i|c_\alpha|j\rangle \langle j|c_\beta^\dagger|i\rangle \\ &= \left\langle c_\alpha \frac{1}{i\omega_\nu + \frac{1}{\hbar}\mu - \frac{1}{\hbar}L} c_\beta^\dagger \right\rangle_{th} - \left\langle c_\beta^\dagger \frac{1}{-i\omega_\nu - \frac{1}{\hbar}\mu - \frac{1}{\hbar}L} c_\alpha \right\rangle_{th} \end{aligned} \quad (9)$$

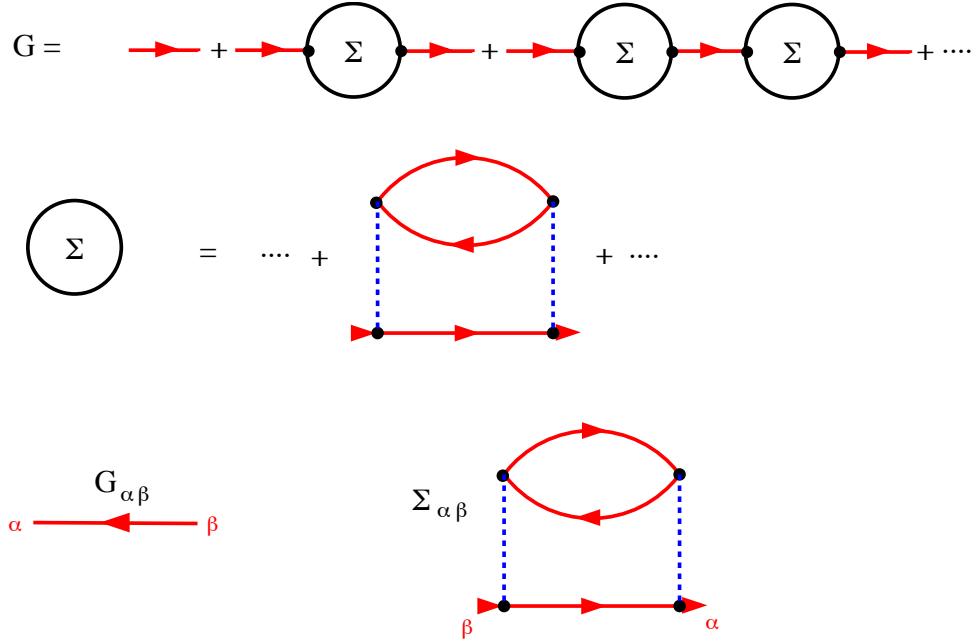


Fig. 1: Top: Graphical representation of the Dyson equation. Middle: Self-energy diagrams have two open ends. Bottom: The convention for the representation of the Green's function implies the labeling of the entry points of the self-energy.

where the Liouville operator L is defined by $LX = [H, X]$. When viewed as function of a complex variable z , all elements of $\mathbf{G}(z)$ are analytic in the complex z -plane except for the real axis, where there are poles at $z = (E_j - E_i - \mu)/\hbar$. It is this property on which the usefulness of the imaginary-time Green's function is based: its Fourier transform $\mathbf{G}(z)$ can be analytically continued to a line infinitesimally above the real axis and then gives the Fourier transform of the retarded real-time Green's function – from which single-particle spectral function i.e. the photoemission and inverse photoemission spectrum of a system can be obtained. For this reason the Fourier transform (9) is often called ‘the’ Green's function and when we speak of the Green's function in the following we always mean (9). The first line in (9) is the *Lehmann representation* of the Green's function.

It is shown in textbooks of field theory [7–9] that the imaginary-time Green's function can be expanded in Feynman diagrams and the self-energy $\Sigma(\mathbf{k}, i\omega_\nu)$ can be introduced in the standard way, see Figure 1. The self-energy can be expanded in diagrams which have two ‘entry points’ an incoming and an outgoing one, see Figure 1. Following FW [8], we represent the Green's function $G_{\alpha\beta}$ by a directed line with an arrow running $\beta \rightarrow \alpha$ (the reason is that it is the *creation operator* that has the index β , see (7)). In the Dyson equation the orbital indices of the Green's function and the self-energy must take the form of consecutive matrix products, e.g. $G_{\delta\alpha}^0 \Sigma_{\alpha\beta} G_{\beta\gamma}^0$ – otherwise the summation of the geometric series would not be possible. It follows that the element $\Sigma_{\alpha\beta}$ must have the label α on the outgoing entry and the label β on the incoming one, see Figure 1. This will be of some importance later on.

Note that the real time Green's function at finite temperature does not allow for a Feynman diagram expansion – this is why the digression of calculating the imaginary time Green's function

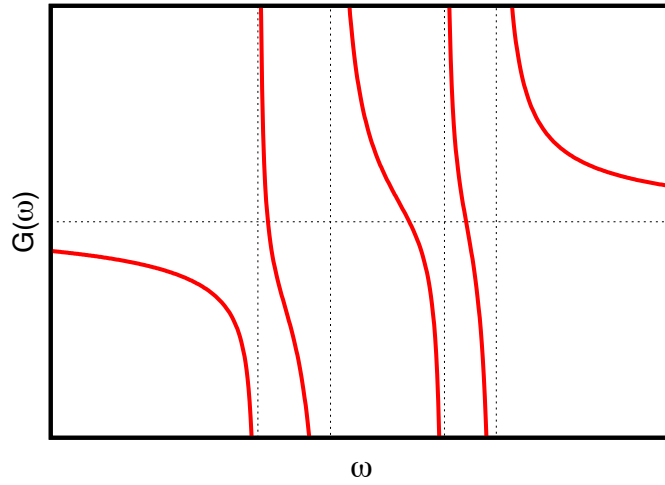


Fig. 2: Real part of the Green's function $G(\omega)$ for real ω . The dashed vertical lines give the position of the poles, ω_i .

and analytically continuing its Fourier transform is necessary. It follows from the diagrammatic expansion that the Green's-function obeys the Dyson equation (see (26.5) of FW)

$$\begin{aligned} \left(i\omega_\nu + \mu/\hbar - \mathbf{t}(\mathbf{k})/\hbar - \Sigma(\mathbf{k}, i\omega_\nu) \right) G(\mathbf{k}, i\omega_\nu) &= 1 \\ \left(-\partial_\tau + \mu/\hbar - \mathbf{t}(\mathbf{k})/\hbar \right) G(\mathbf{k}, \tau) - \int_0^{\beta\hbar} \Sigma(\mathbf{k}, \tau - \tau') G(\mathbf{k}, \tau') d\tau' &= \delta(\tau). \end{aligned} \quad (10)$$

where the second equation is the Fourier-transform of the first and FW (25.21) was used.

Let us finally briefly discuss the analytic structure of the Green's function and self-energy. For simplicity we specialize to a single band and assume that the z -component of spin is a good quantum number so that the Green's function is a scalar. It can be seen from (9) that the Fourier transform of the Green's function has the form

$$G(\omega) = \sum_i \frac{\alpha_i^2}{\omega - \omega_i}$$

where α_i and ω_i are real numbers. It has poles on the real axis and the real part of $G(\omega)$ looks like in Figure 2. This shows that in between any two successive poles ω_i and ω_{i+1} the Green's function crosses zero with a negative slope

$$G(\omega) \approx -\beta_i^2(\omega - \zeta_i).$$

Near the crossing point ζ_i we thus have

$$\begin{aligned} \Sigma(\omega) &= -G^{-1}(\omega) + \omega + \mu - t_{\mathbf{k}} \\ &= \frac{\sigma_i}{\omega - \zeta_i} + \dots \end{aligned}$$

where $\sigma_i = 1/\beta_i^2$. The self-energy thus has poles on the real axis as well, and these poles are 'sandwiched' in between the poles of the Green's function. Luttinger has shown [12] that $\Sigma(\omega)$ is essentially determined by these poles and their residua in that it can be written as

$$\Sigma(\omega) = \eta + \sum_i \frac{\sigma_i}{\omega - \zeta_i} \quad (11)$$

with a real constant η .

3 Proof of the theorem by Luttinger and Ward

3.1 Statement of the theorem

The grand canonical potential $\Omega(T, \mu)$ contains all thermodynamical information about a system at fixed temperature T and chemical potential μ . It is defined as the logarithm of the grand partition function

$$\begin{aligned}\Omega &= -\frac{1}{\beta} \ln(Z), \\ Z &= \sum_i e^{-\beta(E_i - \mu N_i)},\end{aligned}$$

where the sum is over all eigenstates of the system with energy E_i and particle number N_i . The latter can indeed be evaluated for noninteracting particles and in this way one obtains for example the grand canonical potential of noninteracting Bloch electrons

$$\Omega_{\text{non-int}} = -\frac{1}{\beta} \sum_{n=1}^{2n_{\text{orb}}} \sum_{\mathbf{k}} \ln(1 + e^{-\beta(E_n(\mathbf{k}) - \mu)}). \quad (12)$$

As shown in textbooks of statistical mechanics, expression (12) allows one to derive the complete thermodynamics of metals. However, it is in general not possible to evaluate the grand partition function for a system of *interacting particles* of macroscopic size.

Luttinger and Ward, however, derived a relation for the grand canonical potential of interacting fermions [1]. More precisely, they considered the following quantity

$$\tilde{\Omega} = -\lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} \left(\ln \det(-\mathbf{G}^{-1}(\mathbf{k}, i\omega_\nu)) + \text{Tr}(\mathbf{G}(\mathbf{k}, i\omega_\nu) \boldsymbol{\Sigma}(\mathbf{k}, i\omega_\nu)) \right) + \Phi[\mathbf{G}]. \quad (13)$$

Here \sum_ν denotes summation over the Fermionic Matsubara frequencies and $\Phi[\mathbf{G}]$ is the so-called Luttinger-Ward functional which is defined as a sum over closed, linked Feynman diagrams (the precise definition will be discussed below). The important point here is that a closed Feynman diagram is simply a number, so that $\Phi[\mathbf{G}]$ indeed assigns a (real) number to each possible Green's function \mathbf{G} . Regarding the logarithm of the determinant in (13), we recall that the determinant of a matrix is given by the product of its eigenvalues (the matrix need not be Hermitian for this to be true), so the logarithm of the determinant is the sum of the logarithms of the (complex) eigenvalues of $-\mathbf{G}^{-1}$.

In the following, we want to show that in fact $\tilde{\Omega} = \Omega$, the true grand canonical potential and thereby follow the original proof by Luttinger and Ward. The basic idea is to multiply the interaction part of the Hamiltonian, (2), by a scale factor, $H_1 \rightarrow \lambda H_1$, then show $\tilde{\Omega} = \Omega$ for $\lambda = 0$ – i.e. the noninteracting limit – and next show that $\partial_\lambda \tilde{\Omega} = \partial_\lambda \Omega$. Obviously, this proves the identity of the two expressions for any λ .

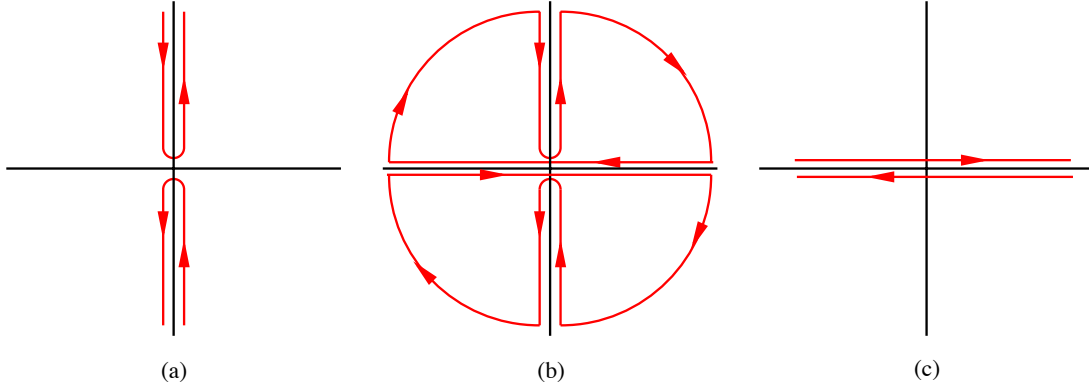


Fig. 3: (a) Integration contour \mathcal{C} used in (15). Since the integrals along the two contours in (b) are zero and the contributions from the circular arcs vanish, the integral along the contour in (a) is equal to that over the contour \mathcal{C}' in (c).

3.2 The case $\lambda = 0$

In this limit, $\Sigma = 0$ and $\Phi[\mathbf{G}] = 0$ (the latter property follows because all interaction lines in all diagrams are zero) so that only the first term in (13) remains and

$$\begin{aligned} \mathbf{G}^{-1}(\mathbf{k}, \omega) &= \omega + (\mu - \mathbf{t}(\mathbf{k})) / \hbar, \\ \ln \det(-\mathbf{G}^{-1}(\mathbf{k}, \omega)) &= \sum_{n=1}^{2n_{orb}} \ln \left(-\omega - (\mu - E_n(\mathbf{k})) / \hbar \right). \end{aligned} \quad (14)$$

We now replace the sum over Matsubara frequencies by a contour integration, a standard trick used in field theory (see e.g. section 25 of FW) and obtain

$$-\frac{1}{\beta} \sum_{\nu} e^{i\omega_{\nu}\eta} \ln \det(-\mathbf{G}^{-1}(\mathbf{k}, i\omega_{\nu})) = \frac{\hbar}{2\pi i} \oint_{\mathcal{C}} d\omega f(\omega) e^{\omega\eta} \ln \det(-\mathbf{G}^{-1}(\mathbf{k}, \omega)), \quad (15)$$

where

$$f(\omega) = \frac{1}{e^{\beta\hbar\omega} + 1},$$

is the Fermi function and the contour \mathcal{C} encircles the imaginary axis in counterclockwise fashion, see Figure 3a. Next we note that the integrals along the two clover-shaped contours in Figure 3b are zero, *provided* the integrand is analytic in the interior of the two curves. Since the Fermi function has all of its poles along the imaginary axis, which is outside of the curves in 3b, we only need to consider possible singularities of $\ln \det(-\mathbf{G}^{-1}(\mathbf{k}, \omega))$. In principle, the complex logarithm has a branch-cut along the negative real axis which could be problematic. However, a quick glance at (14) shows that as long as ω has a nonvanishing imaginary part, the argument of the logarithm can never be purely real. Singularities of the logarithm thus occur only on the real axis, which also is exterior to the contours 3b. The integral along these contours therefore is indeed zero. Next, Jordan's lemma can be invoked to establish that the integral along the large semicircles vanishes. Here, the Fermi function $f(\omega)$ guarantees that the

contribution from the semicircle with $\Re(\omega) > 0$ vanishes, whereas the factor $e^{\omega\eta}$ does the same for the semicircle with $\Re(\omega) < 0$. It follows that the integral along the contour \mathcal{C} in Figure 3a is equal to that along the contour \mathcal{C}' in 3c (note the inverted direction of the curves in 3c as compared to 3b). Next, we insert

$$f(\omega) = -\frac{1}{\beta\hbar} \frac{d}{d\omega} \ln(1 + e^{-\beta\hbar\omega}) \quad (16)$$

and integrate by parts. Thereby the Fermi function and the factor $e^{\eta\omega}$ again make sure that the contributions from $\Re(\omega) = \pm\infty$ vanish and we obtain

$$\begin{aligned} & \frac{1}{\beta} \frac{1}{2\pi i} \oint_{\mathcal{C}'} d\omega \ln(1 + e^{-\beta\hbar\omega}) \frac{d}{d\omega} \left(e^{\eta\omega} \sum_{n=1}^{2n_{orb}} \ln(-\omega + (\mu - E_n(\mathbf{k}))/\hbar) \right) \\ &= \frac{1}{\beta} \frac{1}{2\pi i} \oint_{\mathcal{C}'} d\omega \ln(1 + e^{-\beta\hbar\omega}) e^{\eta\omega} \sum_{n=1}^{2n_{orb}} \frac{1}{\omega + (\mu - E_n(\mathbf{k}))/\hbar} + \mathcal{O}(\eta). \end{aligned}$$

Now we substitute $\hbar\omega \rightarrow z$ and use the theorem of residues (thereby remembering that \mathcal{C}' encircles the poles of the Green's function on the real axis in *clockwise* fashion) and after taking the limit $\eta \rightarrow 0$ obtain the expression (12), which completes the first step of the proof.

3.3 Calculation of $\partial\Omega/\partial\lambda$

To obtain the derivative of the true grand potential Ω with respect to λ we start from the formula

$$\begin{aligned} \lambda \frac{\partial}{\partial\lambda} \Omega(\lambda) &= -\frac{\lambda}{\beta} \frac{\partial}{\partial\lambda} \ln \left(\text{Tr} \left(e^{-\beta(H_0 + \lambda H_1) - \mu N} \right) \right) \\ &= \frac{1}{Z} \text{Tr} \left(\lambda H_1 e^{-\beta(H_0 + \lambda H_1) - \mu N} \right) \\ &= \langle \lambda H_1 \rangle_\lambda \end{aligned}$$

where $\langle \dots \rangle_\lambda$ denotes the thermal average *calculated at interaction strength* λ . The last quantity thus is the expectation value of the interaction Hamiltonian for interaction strength λ . This can be computed by making use of the equation of motion of the Green's function, a procedure found in many textbooks, see e.g. Equation (23.14) of FW. One has

$$\langle \lambda H_1 \rangle_\lambda = -\frac{1}{2} \lim_{\tau \rightarrow 0^-} \sum_{\mathbf{k}} \text{Tr} \left(\hbar \frac{\partial}{\partial\tau} - \mu + \mathbf{t}(\mathbf{k}) \right) \mathbf{G}_\lambda(\mathbf{k}, \tau),$$

where the subscript λ on the Green's function implies that this is the exact Green's function for interaction strength λ . Next we recall the Dyson equation (10), which holds for any λ

$$(-\partial_\tau + (\mu - \mathbf{t}(\mathbf{k}))/\hbar) \mathbf{G}(\mathbf{k}, \tau) - \int_0^{\beta\hbar} d\tau' \boldsymbol{\Sigma}(\mathbf{k}, \tau - \tau') \mathbf{G}(\mathbf{k}, \tau') = \delta(\tau).$$

Since $\delta(\tau < 0) = 0$ we have $\lim_{\tau \rightarrow 0^-} \delta(\tau) = 0$ and obtain

$$\begin{aligned} \lambda \frac{\partial}{\partial\lambda} \Omega(\lambda) &= \frac{\hbar}{2} \lim_{\tau \rightarrow 0^-} \sum_{\mathbf{k}} \int_0^{\beta\hbar} d\tau' \text{Tr}(\boldsymbol{\Sigma}_\lambda(\mathbf{k}, \tau - \tau') \mathbf{G}_\lambda(\mathbf{k}, \tau')) \\ &= \frac{1}{2\beta} \sum_{\mathbf{k}, \nu} \text{Tr} \boldsymbol{\Sigma}_\lambda(\mathbf{k}, i\omega_\nu) \mathbf{G}_\lambda(\mathbf{k}, i\omega_\nu). \end{aligned} \quad (17)$$

3.4 Definition and properties of the Luttinger-Ward functional

As already mentioned, the Luttinger-Ward functional $\Phi[\mathbf{G}]$ is defined as a sum over infinitely many Feynman diagrams with certain properties. The diagrams which contribute are *closed*, which means they have no external lines. They are moreover *connected*, which means that they cannot be decomposed into sub-diagrams that are not connected by lines. And finally, only *skeleton diagrams* are taken into account in the Luttinger-Ward functional. A skeleton diagram is a diagram where no Green's function line contains a self-energy insertion. In other words, it is impossible to draw a box around any part of the diagram so that only two Green's function lines cross the box.

At this point we need to discuss an important property of the skeleton diagrams. Let us consider a self-energy diagram. It contains one Green's function line from the entry-point to the exit-point, and a number of Green's function loops. Starting from the entry-point we may follow the Green's function line and draw a circle around each self-energy insertion that we encounter until we reach the exit point. This procedure will eliminate a number of loops, that means enclose them in a self-energy insertion. Then, we continue along the first interaction line which is not eliminated until we reach a Fermion loop that is not yet eliminated. We follow the Green's function line along this loop and again draw a circle around each self-energy insertion. We proceed to the next interaction line that has not yet been eliminated and so on. We end up with a diagram in which all self-energy insertions are inside circles. Replacing the circles by straight lines we obviously obtain a skeleton-diagram for the self-energy. It is easy to see that the skeleton diagram to which a given self-energy diagram is reduced is unique. All self-energy diagrams thus can be grouped into classes such that all members of one class can be reduced to the same skeleton diagram. Conversely, all members of one class can be obtained by starting out from the skeleton-diagram and inserting the full Green's function for each Green's function line in the diagram, which we write as

$$\Sigma(\mathbf{k}, \omega) = \sum_n \Sigma^{(s,n)}(\mathbf{k}, \omega). \quad (18)$$

Here $\Sigma^{(s,n)}$ denotes the set of all n^{th} order skeleton diagrams (i.e. diagrams with n interaction lines) with the Green's function lines replaced by the full Green's function.

Having defined the diagrams contributing to $\Phi[\mathbf{G}]$ each diagram is now translated into a multiple sum according to the standard Feynman rules for the imaginary-time Green's function in momentum space (see section 25 of FW). However, there is one crucial difference: whereas in a standard Feynman diagram a Green's function line corresponds to a factor $\mathbf{G}^0(\mathbf{k}, \omega)$ (the noninteracting Green's function), in the Luttinger-Ward functional we replace $\mathbf{G}^0(\mathbf{k}, \omega) \rightarrow \mathbf{G}(\mathbf{k}, \omega)$ where $\mathbf{G}(\mathbf{k}, \omega)$ is the argument of the functional $\Phi[\mathbf{G}]$. As an example, the expression corresponding to the diagram in Figure 4 is

$$\left(\frac{-1}{\beta \hbar^2 N}\right)^2 (-1)^2 \sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}} \sum_{\alpha, \beta, \gamma, \delta} \sum_{\alpha_1, \beta_1, \gamma_1, \delta_1} \sum_{\nu, \nu', \mu} V_{\alpha, \beta, \delta, \gamma}(\mathbf{k}, \mathbf{k}', \mathbf{q}) V_{\delta_1, \gamma_1, \alpha_1, \beta_1}(\mathbf{k} + \mathbf{q}, \mathbf{k}' - \mathbf{q}, -\mathbf{q}) \\ \times G_{\alpha_1, \alpha}(\mathbf{k} + \mathbf{q}, i\omega_\nu + \omega_\mu) G_{\delta, \delta_1}(\mathbf{k}, i\omega_\nu) G_{\beta_1, \beta}(\mathbf{k}' - \mathbf{q}, \omega_{\nu'} - \omega_\mu) G_{\gamma, \gamma_1}(\mathbf{k}', \omega_{\nu'}) . \quad (19)$$

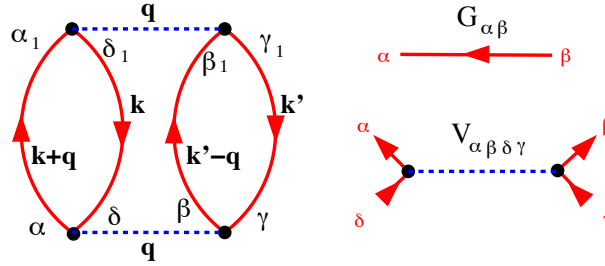


Fig. 4: Left: A diagram contributing to the Luttinger-Ward functional. Right: the elements of the diagram.

The Luttinger-Ward functional $\Phi[\mathbf{G}]$ thus consists of an infinite sum of multiple sums which involve the interaction matrix elements V of the Hamiltonian (4) and the function \mathbf{G} for which the functional is to be evaluated.

Let us briefly discuss the scaling with system size, N . By the Feynman rules an n^{th} order diagram has the prefactor $(1/N)^n$. On the other hand, there are n interaction lines, and $2n$ Green's function lines, so that there are $3n$ momenta. The n interaction lines give rise to $2n$ momentum conservation conditions, one for each end of a line. However, in a *closed* diagram one of these momentum conservation conditions is fulfilled trivially so that there remain $n + 1$ momenta to be summed over (see the above example). Each sum runs over N momenta so that the total diagram is of order N – as it has to be because Ω is an extensive quantity.

In addition to the factors originating from the Feynman rules, each diagram is multiplied by $-1/(\beta S)$ where the positive integer S is the *symmetry factor* of the diagram. A very detailed discussion of these symmetry factors is given in section 2.3 of Negele-Orland [9]. The definition is as follows: first, the diagram is drawn such that all interaction lines are in x -direction. The n interaction lines of a diagram are labeled by integers $i \in \{1 \dots n\}$ and the ends of each interaction line are labeled by R and L (for ‘right end’ and ‘left end’), see Figure 5a. Any Green's function line in the diagram now can be labeled by the ends of the interaction lines where it departs and where it ends: $(i, S_1) \rightarrow (j, S_2)$ with $i, j \in \{1 \dots n\}$ and $S_1, S_2 \in \{R, L\}$. Obviously, the diagram is characterized completely by the $2n$ ‘directed quadruples’ $(i, S_1) \rightarrow (j, S_2)$. Then, we consider the following operations on the diagrams: a) any permutation of the indices i , b) exchange of the labels R and L on an arbitrary number of interaction lines, c) any combination of a permutation followed by label exchanges. Such an operation obviously changes the quadruples which characterize the connectivity of the diagram: $[(i, S_1) \rightarrow (j, S_2)] \rightarrow [(i', S'_1) \rightarrow (j', S'_2)]$. The symmetry factor of a diagram then is the number of symmetry operations – including identity – where the new labels $(i', S'_1) \rightarrow (j', S'_2)$ are a permutation of the old ones, $(i, S_1) \rightarrow (j, S_2)$ (Negele-Orland then call the transformed diagram a deformation of the first one). As an example, consider the diagram in Figure 5a. Label exchange on, say, the interaction line 2 leads to the diagram shown in 5b, which however is not a deformation of the original diagram. This can be seen by considering e.g. the line connecting the R -end of 1 and the R -end of 2. In 5a this line would have the label $(2, R) \rightarrow (1, R)$, whereas it would be $(1, R) \rightarrow (2, R)$ in 5b. This means that the direction of momentum flow along the

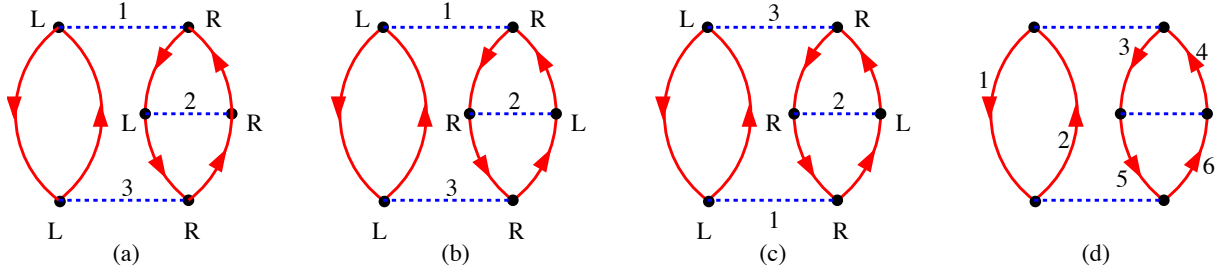


Fig. 5: Determination of the symmetry factor S for a diagram.

line would be reversed. On the other hand, the permutation of the labels 1 and 3 followed by label exchange on interaction line 2 leads to the diagram 5c which indeed is a deformation of the original diagram. In Figure 5d the Green's function lines are numbered by $1 \rightarrow 6$ and Table 1 gives the quadruples corresponding to these lines in Figures 5a and 5c. Obviously the two sets of quadruples are a permutation of each other. It turns out that this is the only symmetry operation which leaves the diagram invariant, so that, taking into account the identity operation, the diagram has $S = 2$. Since a symmetry operation corresponds to a permutation of the quadruples $(i, S_1) \rightarrow (j, S_2)$ that characterize the individual Green's function lines in a diagram, it defines a mapping between these lines whereby each line is mapped onto the one that gets its label. For example, from Table 1 one reads off the corresponding mapping for the operation connecting 5a and 5c:

1	2	3	4	5	6
2	1	6	5	4	3

If two Green's function lines i and j are mapped onto each other the lines are equivalent in the sense that the diagram could be deformed such that the deformed diagram is precisely the same as the original one but with line j now taking the place of line i and vice versa.

Let us now assume that a diagram has the symmetry factor S . This means that all Green's function lines can be grouped into disjunct classes such that the lines belonging to one class are mapped onto each other by one of the S symmetry operations. For example, the diagram in 5 has the classes $(1, 2)$, $(3, 6)$ and $(4, 5)$. Since a diagram with n interaction lines has $2n$ Green's function lines the number of classes is $2n/S$ which will be of importance later on.

Line	5(a)	5(c)
1	(1,L)→(3,L)	(3,L)→(1,L)
2	(3,L)→(1,L)	(1,L)→(3,L)
3	(1,R)→(2,L)	(3,R)→(2,R)
4	(2,R)→(1,R)	(2,L)→(3,R)
5	(2,L)→(3,R)	(2,R)→(1,R)
6	(3,R)→(2,R)	(1,R)→(2,L)

Table 1: Quadruples describing the connectivity of the diagrams Figure 5a and Figure 5c. The numbers of the Green's function lines are given in Figure 5d.

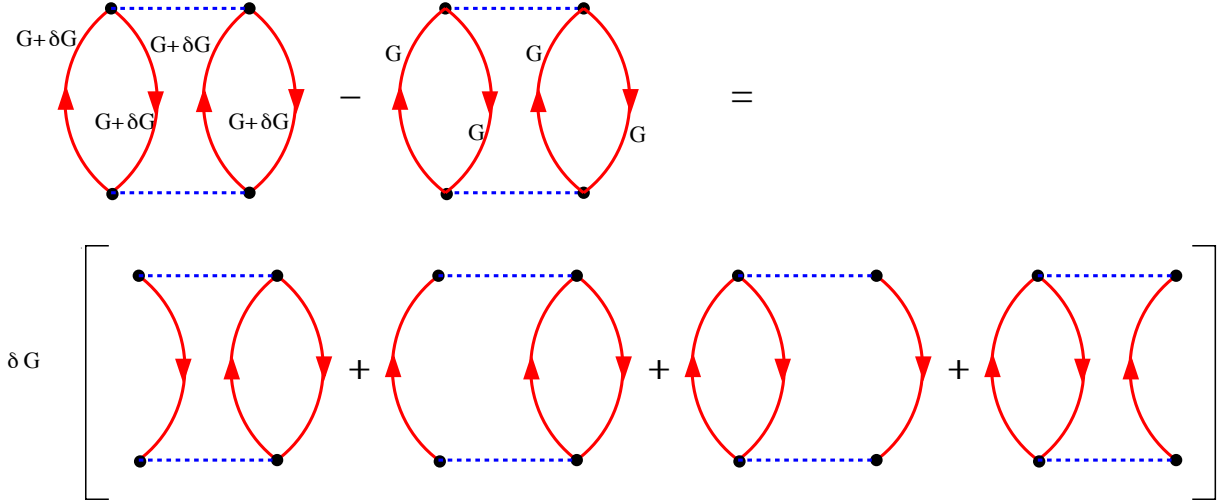


Fig. 6: Variation of G implies opening the lines of a Feynman diagram.

Next, we want to see the meaning of this definition. In fact, the Luttinger-Ward functional is the *generating functional* of the self-energy, or, more precisely:

$$\frac{\partial \Phi}{\partial G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)} = \frac{1}{\beta} \Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu). \quad (20)$$

To see this, consider an infinitesimal change $G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \rightarrow G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) + \delta G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)$ as in Figure 6. The initial diagrams correspond to multiple sums over products of Green's functions where all internal frequencies, momenta, and orbital indices are summed over, subject to the condition of energy/momentum conservation at each interaction vertex, see (19). The first-order change then also can be viewed as a sum of diagrams but with a single missing line, which corresponds to the variation δG that has been factored out. Another way to state this is to say that differentiating with respect to an element of G amounts to successively opening each of the lines in the initial closed diagram and summing the remaining diagrams. These remaining diagrams obviously look like self-energy diagrams in that they have two entry points. We now need to show, however, that the diagrams not only look like possible contributions to the skeleton diagram expansion of the self-energy, but that they come with exactly the right numerical prefactors. At this point, the additional prefactors of $-1/\beta S$ turn out to be crucial.

We first note that the momentum and frequency which flow into/out-of the diagram are fixed by the momentum and frequency of δG . Regarding the orbital indices, we recall that $G_{\alpha\beta}$ corresponds to a directed line $\beta \rightarrow \alpha$. The resulting self-energy-like diagrams therefore all have the matrix index α on their incoming entry and β on their outgoing entry and, comparing with Fig. 4, we see that this assignment of indices corresponds to $\Sigma_{\beta\alpha}$. Moreover, all internal momenta, frequencies and matrix indices in the remaining diagrams are summed over – subject to the condition of frequency and momentum conservation at the interaction lines – as would be the case in the true self-energy diagrams. Second, the order n of a diagram, i.e. the number of interaction lines, is not changed by opening a Green's function line, so that the prefactor $(-1/\beta \hbar^2 N)^n$ of the closed diagram is also the correct prefactor for the resulting self-energy diagram. Third, opening a Green's function line reduces the number of closed fermion loops by 1

and the factor (-1) in $-1/\beta S$ takes care of this. Lastly, we need to discuss the symmetry factor S . Let us consider a diagram with n interaction lines, which accordingly has $2n$ Green's function lines and moreover assume that the diagram has the symmetry factor S . As we saw above, the $2n$ Green's function lines can be divided into classes of S members which are mapped into each other by the symmetry operations, and the number of these classes is $2n/S$. A symmetry operation maps a Green's function line i onto an equivalent one j , so it is possible to deform the diagram such that it looks exactly the same as the original one but with line j in place of line i . This means, however, that 'opening' the line i also gives *exactly* the same self-energy diagram as opening line j . Accordingly, from the single closed diagram of degree n with symmetry factor S we obtain $2n/S$ different skeleton diagrams for the self-energy, and each is produced S times, see also Figure 7. This factor of S , however, precisely cancels the prefactor $1/S$. It follows that each skeleton-diagram for the self-energy is produced with the same prefactor $1/\beta$. Differentiating $\Phi[\mathbf{G}]$ with respect to $G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)$ thus gives $1/\beta$ times the sum of all skeleton diagrams for $\Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu)$, with the noninteracting Green's function replaced by the full one, and this is exactly $\Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu)$ itself, see (18), proving (20).

We have just seen that all skeleton-diagrams for the self-energy can be obtained by differentiating the Luttinger-Ward functional with respect to \mathbf{G} , whereby the differentiation corresponds to opening one line in a closed diagram. We then may ask if this operation can be reversed, namely if the Luttinger-Ward functional can be obtained by starting from the skeleton-diagram expansion of the self-energy and closing the diagrams by reconnecting the entry-points of the self-energy by a Green's function. More precisely, we consider an expression of the form

$$\frac{1}{\beta} \sum_{\nu, \mathbf{k}} \sum_{\alpha, \beta} \mathbf{G}_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \Sigma_{\beta\alpha}^{(s,n)}(\mathbf{k}, i\omega_\nu). \quad (21)$$

We have seen that an n^{th} order diagram contributing to $\Phi[\mathbf{G}]$ with symmetry factor S produces $2n/S$ different skeleton-self-energy diagrams, and each of them S times and with a factor of (-1) , so that the remaining prefactor was $1/\beta$. Upon closing the fermion line again, according to (21), each of these diagrams produces the original closed diagram (it is easy to see that for each self-energy diagram there is exactly one closed diagram from which it can be obtained). Since there are $2n/S$ self-energy diagrams originating from the original closed diagram the latter is produced $2n/S$ times and thus has the additional prefactor $-2n/S\beta$, where the factor of (-1) is due to the additional fermion loop in the closed diagram. In the expansion of $\Phi[\mathbf{G}]$, however, the diagram would have had the prefactor $-1/S\beta$, or, put another way, closing the sum of all n^{th} order skeleton diagrams for Σ according to (21) produces the n^{th} order contribution to $\Phi[\mathbf{G}]$ with an additional prefactor of $2n$ so that

$$\Phi^{(n)} = \frac{1}{2n\beta} \sum_{\nu, \mathbf{k}} \text{Tr} \mathbf{G}(\mathbf{k}, i\omega_\nu) \Sigma^{(s,n)}(\mathbf{k}, i\omega_\nu). \quad (22)$$

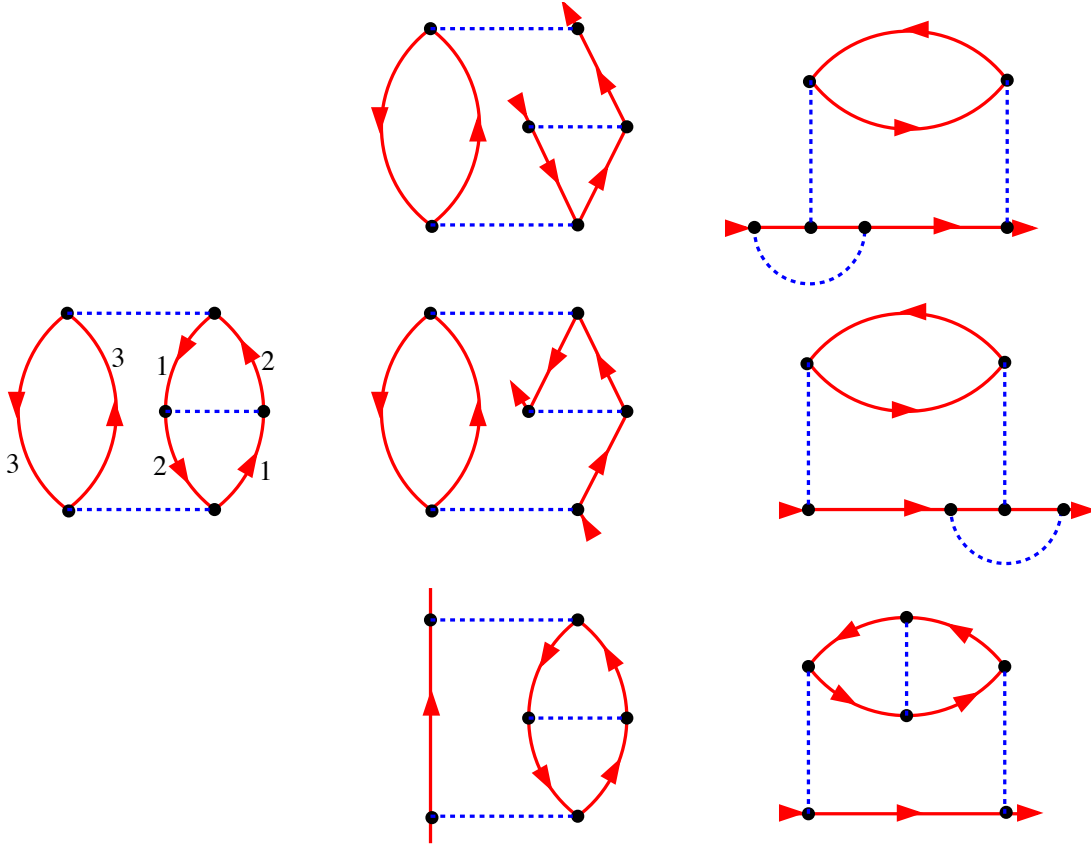


Fig. 7: The diagram on the left has $n = 3$ and $S = 2$ and accordingly 3 classes of symmetry-equivalent Green's function lines. The lines are labeled by the number of the classes, compare Figure (5) and Table 1. Successively opening the lines of the diagram produces the three different self-energy diagrams in the center column and each of them is produced $S = 2$ times. The right column shows the diagrams redrawn to more look like self-energy diagrams.

3.5 Calculation of $\partial\tilde{\Omega}/\partial\lambda$

We proceed to the the final step of the proof and compute $\partial\tilde{\Omega}/\partial\lambda$. If we vary the interaction strength λ there are two places in the expression $\tilde{\Omega}$ in (13) where this makes itself felt. Namely, the self-energy Σ will change, and moreover the interaction matrix elements V in the Luttinger-Ward functional (compare 19) that have a prefactor of λ will also contribute to the variation. Let us first consider the variation of Σ and compute

$$\frac{\partial\tilde{\Omega}}{\partial\Sigma_{\alpha,\beta}(\mathbf{k}, i\omega_\nu)}.$$

There are three terms in (13) and we consider them one after the other. The first two terms involve a sum over momentum and frequency and obviously only those terms with momentum \mathbf{k} and frequency ω_ν will contribute. Accordingly, in the following equations we omit the

arguments $(\mathbf{k}, i\omega_\nu)$ for brevity. Then we find by using the chain rule for differentiation

$$\begin{aligned} \frac{\partial}{\partial \Sigma_{\alpha,\beta}} \left(-\frac{1}{\beta} \ln \det (-\mathbf{G}^{-1}) \right) &= -\frac{1}{\beta} \sum_{\mu,\nu} \left(\frac{\partial}{\partial (-G_{\mu,\nu}^{-1})} \ln \det (-\mathbf{G}^{-1}) \right) \frac{\partial (-G_{\mu,\nu}^{-1})}{\partial \Sigma_{\alpha,\beta}} \\ &= -\frac{1}{\beta} \sum_{\mu,\nu} (-G_{\nu,\mu}) \delta_{\mu,\alpha} \delta_{\nu,\beta} \\ &= \frac{1}{\beta} G_{\beta,\alpha} . \end{aligned}$$

In going to the 2nd line we used the identity from the Appendix and the Dyson equation

$$-\mathbf{G}^{-1} = -\omega - \mu/\hbar + \Sigma$$

from which it follows that

$$\frac{\partial (-G_{\mu,\nu}^{-1})}{\partial \Sigma_{\alpha,\beta}} = \delta_{\mu,\alpha} \delta_{\nu,\beta} .$$

We proceed to the second term:

$$\frac{\partial}{\partial \Sigma_{\alpha,\beta}} \left(-\frac{1}{\beta} \text{Tr} \Sigma \mathbf{G} \right) = \frac{\partial}{\partial \Sigma_{\alpha,\beta}} \left(-\frac{1}{\beta} \sum_{\mu,\nu} \Sigma_{\nu,\mu} G_{\mu,\nu} \right) = -\frac{1}{\beta} \left(G_{\beta,\alpha} + \sum_{\mu,\nu} \Sigma_{\nu,\mu} \frac{\partial G_{\mu,\nu}}{\partial \Sigma_{\alpha,\beta}} \right) .$$

Lastly we consider the Luttinger-Ward functional. Using again the chain rule we find

$$\frac{\partial \Phi[G]}{\partial \Sigma_{\alpha,\beta}} = \sum_{\mu,\nu} \frac{\partial \Phi[G]}{\partial G_{\mu,\nu}} \frac{\partial G_{\mu,\nu}}{\partial \Sigma_{\alpha,\beta}} = \frac{1}{\beta} \sum_{\mu,\nu} \Sigma_{\nu,\mu} \frac{\partial G_{\mu,\nu}}{\partial \Sigma_{\alpha,\beta}} .$$

Adding up the three terms we thus obtain the important result

$$\frac{\partial \tilde{\Omega}}{\partial \Sigma_{\alpha,\beta}(\mathbf{k}, i\omega_\nu)} = 0 . \quad (23)$$

In other words: the expression $\tilde{\Omega}$, which will be seen to be equal to the grand potential Ω in a moment, is *stationary with respect to variations of the self-energy*. This is the stationarity condition for Σ which is the basis of the VCA.

First, however, we have to complete the proof and evaluate $\lambda \frac{\partial}{\partial \lambda} \tilde{\Omega}(\lambda)$. Since there is no variation of $\tilde{\Omega}$ due to a variation of Σ , the only remaining source of variation are the interaction lines in the Luttinger-Ward functional. Namely any n^{th} order diagram has the prefactor of λ^n so that

$$\lambda \frac{\partial}{\partial \lambda} \Phi^{(n)} = n \Phi^{(n)}$$

Using (22) we thus obtain

$$\begin{aligned} \lambda \frac{d\tilde{\Omega}}{d\lambda} &= \sum_n n \Phi^{(n)} = \sum_n \frac{1}{2\beta} \sum_{\nu,\mathbf{k}} \text{Tr} \mathbf{G}_\lambda(\mathbf{k}, i\omega_\nu) \Sigma_\lambda^{(s,n)}(\mathbf{k}, i\omega_\nu) \\ &= \frac{1}{2\beta} \sum_{\nu,\mathbf{k}} \text{Tr} \mathbf{G}_\lambda(\mathbf{k}, i\omega_\nu) \left(\sum_n \Sigma_\lambda^{(s,n)}(\mathbf{k}, i\omega_\nu) \right) \\ &= \frac{1}{2\beta} \sum_{\nu,\mathbf{k}} \text{Tr} \mathbf{G}_\lambda(\mathbf{k}, i\omega_\nu) \Sigma_\lambda(\mathbf{k}, i\omega_\nu) . \end{aligned}$$

Comparing with (17), we see that this is equal to $\lambda \frac{\partial}{\partial \lambda} \Omega(\lambda)$ which completes the proof. Let us summarize the results that we have obtained:

1. The grand canonical potential Ω of an interacting Fermi system is given by eqn. (13).
2. The Luttinger-Ward functional is the generating functional of $\Sigma(\mathbf{k}, i\omega_\nu)$, see eqn. (20).
3. $\Phi[\mathbf{G}]$ depends only on the interaction matrix elements $V_{\alpha\beta\delta\gamma}$ in the Hamiltonian and the Green's function \mathbf{G} which is the argument of the functional.
4. Ω is stationary under variations of $\Sigma(\mathbf{k}, i\omega_\nu)$ see (23).

Looking at the above proof one might worry about the fact that it assumes a continuous evolution of the system with increasing interaction strength λ - whereas we are interested e.g. in Mott-insulators where we have reason to believe that a phase transition occurs as a function of λ . However, Potthoff has recently given a nonperturbative proof of the theorem [6, 13] that means all of the above properties of the grand potential, the Luttinger-Ward functional and the self-energy remain valid in a strongly correlated electron system where a Feynman-diagram expansion of the Green's function and the adiabatic continuity with the noninteracting system can no longer be assumed valid.

4 The variational cluster approximation

In the preceding section we have seen that the grand canonical ensemble of a system is stationary with respect to variations of the self-energy. In order to rewrite Ω as a functional of the self-energy we need to change the argument of the Luttinger-Ward functional from \mathbf{G} to Σ . Since Σ is the derivative of Φ with respect to \mathbf{G} this can be achieved, following [4], by introducing the Legendre-transform of the Luttinger-Ward functional:

$$\begin{aligned} F[\Sigma] &= \Phi[\mathbf{G}[\Sigma]] - \sum_{\mathbf{k}, \nu} \sum_{\alpha, \beta} \frac{\partial \Phi}{\partial G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)} G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \\ &= \Phi[\mathbf{G}[\Sigma]] - \frac{1}{\beta} \sum_{\mathbf{k}, \nu} \sum_{\alpha, \beta} G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu) \\ &= \Phi[\mathbf{G}[\Sigma]] - \frac{1}{\beta} \sum_{\mathbf{k}, \nu} \text{Tr } \mathbf{G}(\mathbf{k}, i\omega_\nu) \Sigma(\mathbf{k}, i\omega_\nu) . \end{aligned}$$

By virtue of being a Legendre transform this new functional obviously satisfies

$$\frac{\partial F}{\partial \Sigma_{\alpha\beta}(\mathbf{k}, i\omega_\nu)} = -\frac{1}{\beta} G_{\beta\alpha}(\mathbf{k}, i\omega_\nu) . \quad (24)$$

Moreover, the second and third term in (13) together are nothing but $F[\Sigma]$, whence

$$\Omega = - \lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} [\ln \det (-i\omega_\nu + (\mathbf{t}(\mathbf{k}) - \mu)/\hbar + \Sigma(\mathbf{k}, i\omega_\nu))] + F[\Sigma] . \quad (25)$$

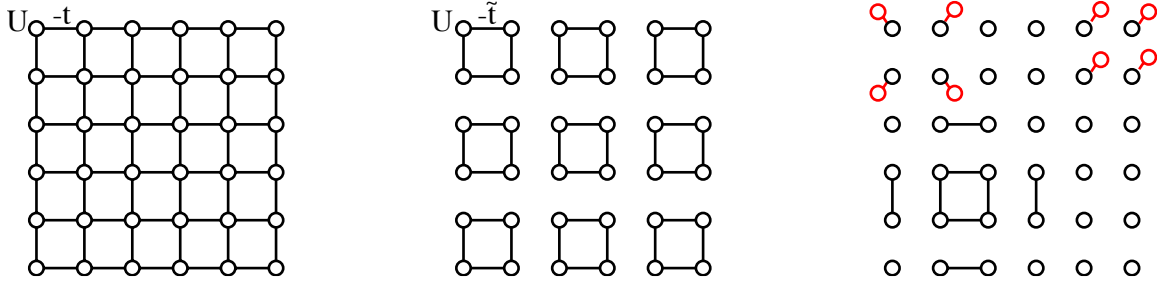


Fig. 8: *Left: The 2D Hubbard model. Center: The 2D Hubbard model partitioned into 2×2 clusters. Right: Partitioning into larger/smaller clusters, with or without additional bath sites and additional hopping integrals. All systems have a different kinetic energy H_0 but exactly the same interaction part H_1 ; accordingly, they all have the same Luttinger Ward functional.*

Here we have used the Dyson equation to replace $-\mathbf{G}^{-1}$ in the first term. This expresses Ω as a functional of Σ , and this functional is known to be stationary with respect to variations of its argument at the exact Σ (this is also easily verified using the identity from the Appendix together with (24)). One might now try and either derive Euler-Lagrange equations or introduce a trial self-energy containing some variational parameters, e.g., of the form (11) with only a certain number of poles (σ_i, ζ_i) retained, and perform the variation with respect to these parameters. Unfortunately this procedure does not work, because the functional $F[\Sigma]$ was defined as the Legendre transform of the Luttinger-Ward functional $\Phi[\mathbf{G}]$, which in turn was defined as a sum over an infinite number of Feynman diagrams and thus is completely impossible to evaluate for a given trial self-energy.

At this point Potthoff's new idea comes into play. For definiteness let us assume that we are interested in a 2D Hubbard model, shown schematically in Figure 8a. Then, we might partition the plane into finite clusters and set the hopping between the clusters to zero, so that they become disconnected, see Fig. 8b. The resulting array of clusters has been termed the *reference system*. The finite clusters also can be decorated in various ways by noninteracting bath orbitals, they can be larger than just 2×2 or contain hopping terms not included in the original Hubbard Hamiltonian, see Fig. 8c. As long as the resulting clusters are not too big, the Fock space of a single cluster has a manageable size and the clusters can be treated by exact diagonalization. This gives us all eigenstates $|i\rangle$ together with their energies E_i and particle numbers N_i . Using these, we may numerically evaluate the grand partition function \check{Z} and obtain the potential $\check{\Omega}$ for a single cluster. In addition, we can calculate the Green's function matrix $\check{\mathbf{G}}(\omega)$ using the Lehmann representation or the Lanczos algorithm [14], invert it and extract the self-energy $\check{\Sigma}(\omega)$. For all of this it is actually sufficient to know all eigenstates with $E_i - \mu N_i$ within a window of $\approx 10 k_B T$ above the minimum value, which can be obtained by the Lanczos algorithm even for clusters of size $\tilde{N} \approx 10 - 20$. Next, we revert to expression (25) and obtain the numerical value of $F[\check{\Sigma}]$:

$$F[\check{\Sigma}] = \check{\Omega} + \lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\nu} e^{i\omega_{\nu}\eta} \left[\ln \det \left(-i\omega_{\nu} + (\check{\mathbf{t}} - \mu)/\hbar + \check{\Sigma}(i\omega_{\nu}) \right) \right], \quad (26)$$

where $\check{\mathbf{t}}$ is the kinetic energy of the cluster, i.e., the matrix $t_{\alpha\beta}$ in (1). This procedure gives

us the *exact* self-energy $\check{\Sigma}(\omega)$ of the cluster together with the *exact* numerical value of the corresponding Luttinger-Ward functional. An important point is that all matrix elements $\check{\Sigma}_{\alpha\beta}$ that have one index α or β on a bath site are zero. This can be seen, e.g., from the diagrammatic expansion of the self-energy. The cluster self-energy $\check{\Sigma}$ therefore has non-vanishing entries only for the correlated sites of the original lattice problem.

At this point the crucial observation by Potthoff comes into play: we have seen above that the Luttinger-Ward functional $\Phi[\mathbf{G}]$ was a sum over Feynman diagrams into which, apart from numerical factors, only two quantities do enter: the interaction matrix elements V of the Hamiltonian and the Green's function \mathbf{G} which is the argument of the functional, see e.g. (19). In our example with the 2D Hubbard model, however, the full 2D Hubbard model and the array of clusters, which may include non-interacting bath sites, differ only in their single-particle terms H_0 but do have exactly the same interaction part H_1 . It follows that the functional $\Phi[\mathbf{G}]$ and hence its Legendre transform $F[\Sigma]$ are identical for the two systems. Since, however, we are able to calculate the self-energy of the cluster and the corresponding value of the Luttinger-Ward functional exactly, we may use these as trial self-energies for the lattice system. In other words, we make the ansatz for the lattice system

$$\begin{aligned}\Omega_{latt} &= - \lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} [\ln \det (-\mathbf{G}'(\mathbf{k}, i\omega_\nu)^{-1})] + F[\check{\Sigma}], \\ \mathbf{G}'(\mathbf{k}, \omega) &= \left(\omega + (\mu - \mathbf{t}(\mathbf{k}))/\hbar - \check{\Sigma}(\mathbf{k}, \omega) \right)^{-1}.\end{aligned}\quad (27)$$

Here $\mathbf{t}(\mathbf{k})$ is now the kinetic energy matrix of the *lattice system* whereas $\check{\Sigma}(\mathbf{k}, \omega)$ is the spatial Fourier transform of the *cluster self-energy* (which may have no \mathbf{k} -dependence at all, depending on the geometry of the reference system). Accordingly, \mathbf{G}' is the approximate Green's function of the lattice system.

Then, how do we perform the variation of the self-energy? The answer is that the single-particle Hamiltonian \check{H}_0 of the cluster used to compute the trial self-energies $\check{\Sigma}$ is completely arbitrary, because the only requirement for the equality of the Luttinger-Ward functionals was the equality of the interaction part H_1 . If we change the single-particle terms of the reference system, i.e. the hopping integrals or site-energies, the self-energy of the cluster will change. The self-energy and its Luttinger-Ward functional thus become functions of the single-particle terms $\check{t}_{\alpha\beta}$ of the reference system: $\Omega_{latt} = \Omega_{latt}(\check{t}_{\alpha\beta})$. Then, we demand that

$$\frac{\partial \Omega_{latt}}{\partial \check{t}_{\alpha\beta}} = 0, \quad (28)$$

which is a condition on the parameters of the reference system, \check{t} , and we denote the solution of (28) by \check{t}^* . The physical interpretation is that the VCA is seeking the best approximation to the self-energy of the lattice-system amongst those functions $\check{\Sigma}(\mathbf{k}, i\omega_\nu)$ that can be represented as the exact self-energies of the reference system for some values of the single-particle parameters $\check{t}_{\alpha\beta}$. After solving (28) we obtain an approximate self-energy $\check{\Sigma}(\mathbf{k}, i\omega_\nu)$ and an approximate value of the grand canonical potential Ω_{latt} . Since Ω_{latt} can be obtained for arbitrary values

of T and μ or other external parameters, thermodynamical quantities such as particle number, entropy or specific heat can be obtained by doing the procedure for different T and μ and differentiating.

As an example we address an interesting property of the VCA. The particle number N_e of any system can be obtained in two different ways (the second is a combination of FW (23.9) and (25.10)):

$$N_e = -\frac{\partial \Omega}{\partial \mu} = \frac{1}{\beta \hbar} \lim_{\eta \rightarrow 0^+} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} \text{Tr } \mathbf{G}'(\mathbf{k}, i\omega_\nu).$$

Since the VCA gives both Ω , and $\mathbf{G}'(\omega)$, it is natural to ask if the two ways of calculating N_e give the same result, and this question has been addressed by Aichhorn *et al.* [15]. We first note that the chemical potential of the reference system has to be the same as that of the physical system. Next, let us assume that we regroup the orbital energies of the cluster, $\tilde{t}_{\alpha\alpha}$, by separating the center of gravity

$$\epsilon = \frac{1}{2n_{orb}} \text{Tr } \tilde{t}$$

and introducing $2n_{orb} - 1$ relative energies $\tilde{t}'_{\alpha\alpha}$ so that $\tilde{t}_{\alpha\alpha} = \tilde{t}'_{\alpha\alpha} + \epsilon$. Since in all calculations for the reference system the chemical potential μ and ϵ only appear in the combination $\mu - \epsilon$, the derivative of any cluster quantity \check{A} with respect to the chemical potential μ obeys

$$\left. \frac{\partial \check{A}}{\partial \epsilon} \right|_{\tilde{t}', \mu} = - \left. \frac{\partial \check{A}}{\partial \mu} \right|_{\tilde{t}', \epsilon}.$$

Next we consider the change of the approximate Ω_{latt} induced by a change of μ . A variation of μ will make itself felt at a variety of places. Looking at (26) and (27) we see that μ appears explicitly in these. Moreover, μ appears in the grand partition function \check{Z} and the Green's function $\check{\mathbf{G}}$ of the reference system, so that the cluster self-energy itself will change with μ . As a consequence of these changes, we have to take into account that \check{t}^* , the solution of (28), will change as well: $\check{t}^* \rightarrow \check{t}^* + \delta \check{t}^*$, so that the situation becomes somewhat complicated. Fortunately enough, the first-order change of Ω_{latt} due to a variation of \check{t}^* is zero; this is exactly the stationarity condition (28). We thus need to consider only the change of Ω_{latt} for fixed parameters \check{t} . Using the last identity in the Appendix we obtain

$$\begin{aligned} - \left. \frac{\partial \Omega_{latt}}{\partial \mu} \right|_{\tilde{t}', \epsilon} &= \lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} \text{Tr} \left(\mathbf{G}'(\mathbf{k}, i\omega_\nu) \left(\frac{1}{\hbar} - \left. \frac{\partial \check{\Sigma}(\mathbf{k}, i\omega_\nu)}{\partial \mu} \right|_{\tilde{t}', \epsilon} \right) \right) - \left. \frac{\partial F[\check{\Sigma}]}{\partial \mu} \right|_{\tilde{t}', \epsilon} \\ &= \lim_{\eta \rightarrow 0^+} \frac{1}{\beta} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} \text{Tr} \left(\mathbf{G}'(\mathbf{k}, i\omega_\nu) \left(\frac{1}{\hbar} + \left. \frac{\partial \check{\Sigma}(\mathbf{k}, i\omega_\nu)}{\partial \epsilon} \right|_{\tilde{t}', \mu} \right) \right) + \left. \frac{\partial F[\check{\Sigma}]}{\partial \epsilon} \right|_{\tilde{t}', \mu} \\ &= \lim_{\eta \rightarrow 0^+} \frac{1}{\beta \hbar} \sum_{\mathbf{k}, \nu} e^{i\omega_\nu \eta} \text{Tr } \mathbf{G}'(\mathbf{k}, i\omega_\nu) - \left. \frac{\partial \Omega_{latt}}{\partial \epsilon} \right|_{\tilde{t}', \mu}. \end{aligned}$$

The presence of the first term in the last line can be understood by noting that μ appears *explicitly* in the approximate cluster Green's function \mathbf{G}' (see (27)) whereas ϵ does not. At this

point we note that if ϵ has been included into the set of cluster parameters which are subject to variation, the last term vanishes (because this is exactly equation (28) for $\check{t}_{\alpha\beta} = \epsilon$) and the two expressions for the particle number indeed give the same result. The VCA thus gives a thermodynamically consistent particle number if and only if the center of gravity of the orbital energies in the reference system is included into the set of parameters to be varied.

To conclude this section we briefly comment on the evaluation of terms like

$$S = -\frac{1}{\beta} \sum_{\nu} e^{i\omega_{\nu}0^+} \ln \det(-\mathbf{G}^{-1}(i\omega_{\nu})),$$

where \mathbf{G} may be either the Green's function of the reference system as in (26) or the approximate Green's function of the lattice system as in (27). The form of this term suggests that we proceed exactly as in the first step of the proof of the Luttinger-Ward theorem, namely to convert the sum over Matsubara-frequencies into a contour-integral, deform the contour using Jordan's lemma as in Figure 3, replace the Fermi function according to (16) and integrate by parts. One obtains

$$S = -\frac{1}{2\pi\beta i} \oint_{C'} d\omega \log(1 + e^{-\beta\omega}) \sum_n \frac{1}{\lambda_n(\omega)} \frac{\partial \lambda_n(\omega)}{\partial \omega}$$

where $\lambda_n(\omega)$ are the eigenvalues of $\mathbf{G}(\omega)$. There are two types of singularities of the integrand in this expression:

1. zeros of an eigenvalue (which corresponds to a singularity of an eigenvalue of Σ) i.e.

$$\lambda(\omega) \approx a_{\nu}(\omega - \zeta_{\nu}) \rightarrow \frac{1}{\lambda(\omega)} \frac{\partial \lambda(\omega)}{\partial \omega} = \frac{1}{\omega - \zeta_{\nu}}$$

2. singularities of an eigenvalue, i.e.

$$\lambda(\omega) \approx \frac{b_{\mu}}{\omega - \eta_{\mu}} \rightarrow \frac{1}{\lambda(\omega)} \frac{\partial \lambda(\omega)}{\partial \omega} = -\frac{1}{\omega - \eta_{\mu}}.$$

In this way we obtain the expression derived by Potthoff [4]:

$$S = -\frac{1}{\beta} \left(\sum_{\mu} \log(1 + e^{-\beta\eta_{\mu}}) - \sum_{\nu} \log(1 + e^{-\beta\zeta_{\nu}}) \right),$$

An alternative is to simply evaluate the contour integral numerically.

5 Applications of the VCA

5.1 Metal-insulator transition in a dimer

As a simple illustration of the procedure we study Potthoff's re-derivation of the phase diagram for the metal-insulator transition in the Hubbard model [16]. We consider a half-filled single-

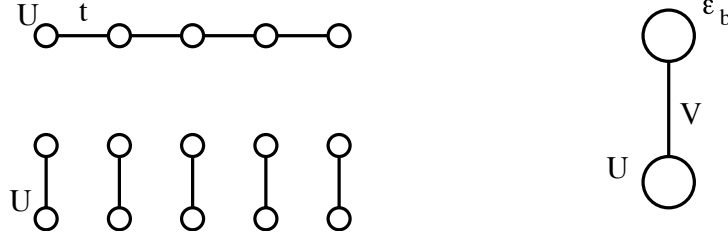


Fig. 9: Left: The physical Hubbard model (top) which is a true infinite lattice system and the reference system which is an array of identical dimers. Right: Schematic representation of the parameters of a single dimer (see the Hamiltonian (29)).

band Hubbard model on a bipartite N -site lattice

$$\begin{aligned}
 H &= \sum_{\mathbf{k}, \sigma} t(\mathbf{k}) c_{\mathbf{k}, \sigma}^\dagger c_{\mathbf{k}, \sigma} + \frac{U}{2} \sum_{i=1}^N (n_i - 1)(n_i - 1) - N \frac{U}{2} \\
 &= \sum_{\mathbf{k}, \sigma} t(\mathbf{k}) c_{\mathbf{k}, \sigma}^\dagger c_{\mathbf{k}, \sigma} + U \sum_{i=1}^N n_{i, \uparrow} n_{i, \downarrow} - \frac{U}{2} \sum_{i=1}^N n_i
 \end{aligned}$$

where $n_i = c_{i, \uparrow}^\dagger c_{i, \uparrow} + c_{i, \downarrow}^\dagger c_{i, \downarrow}$. For simplicity we assume this Hamiltonian has particle-hole symmetry. More precisely, under the transformation $c^\dagger \leftrightarrow c$ we have $n_i - 1 \rightarrow 1 - n_i$ so that the interaction part is invariant, whereas the first term changes sign. If the hopping term connects only sites on different sublattices, which is what we assume, this sign change can be compensated by the gauge transformation $c_{i, \sigma}^\dagger \rightarrow -c_{i, \sigma}^\dagger$ on the sites i of one sublattice. This transformation exchanges photoemission and inverse photoemission spectrum and implies $\mu = U/2$.

For the reference system, Potthoff chose N dimers with one ‘Hubbard-site’ hybridizing with one bath-site, see Figure 9, whereby the Hamiltonian for one dimer reads

$$H - \mu N = -V \sum_{\sigma} (c_{\sigma}^\dagger b_{\sigma} + b_{\sigma}^\dagger c_{\sigma}) + \left(\epsilon_b - \frac{U}{2} \right) \sum_{\sigma} b_{\sigma}^\dagger b_{\sigma} + \frac{U}{2} (n_c - 1)(n_c - 1) - \frac{U}{2}. \quad (29)$$

Here b_{σ}^\dagger creates an electron in a bath site and $n_c = c_{\uparrow}^\dagger c_{\uparrow} + c_{\downarrow}^\dagger c_{\downarrow}$. We have to write $\epsilon_b - U/2$ because $\mu = U/2$. Since we want to generate particle-hole symmetric self-energies we have to impose particle-hole symmetry also in the reference system. The transformation $c^\dagger \leftrightarrow c$, $b^\dagger \leftrightarrow -b$ indeed converts the Hamiltonian into itself *except* for the second term. Setting $\epsilon_b = U/2$, however, eliminates this term and particle-hole symmetry is restored. The only remaining parameter to be varied therefore is V .

The Fock space of the dimer has a dimension of $4 \times 4 = 16$, so all eigenstates can be readily obtained. If we construct basis functions with fixed particle number, spin, and z -component of the spin, the problem in fact can be broken down to diagonalizing 2×2 matrices, i.e. the reference system can be solved analytically. To further simplify the calculations, Potthoff used a semielliptical density of states of width $W = 4$ for the conduction band

$$\rho_0(\epsilon) = \frac{1}{2\pi} \sqrt{4 - \epsilon^2}.$$

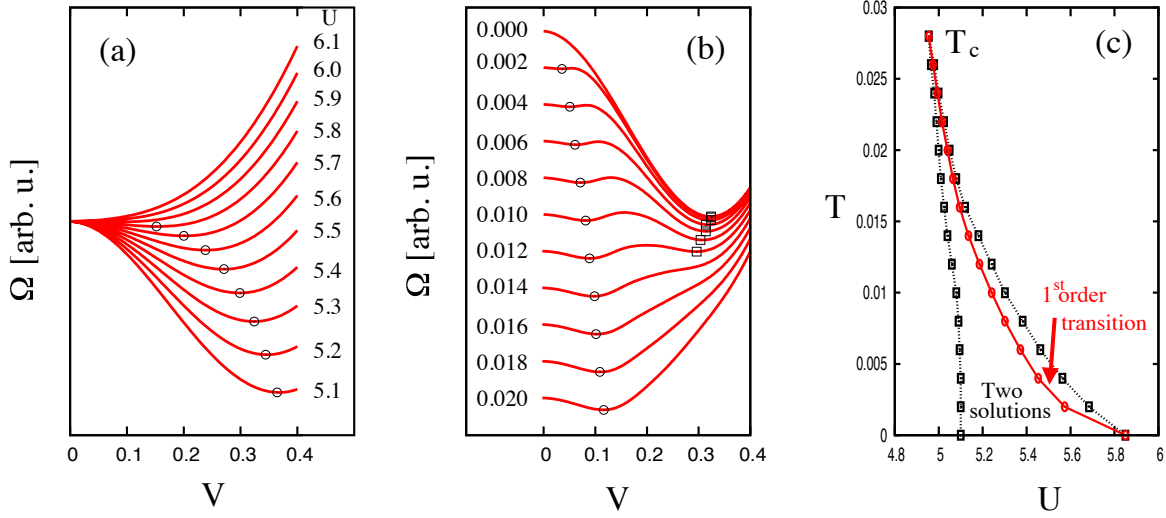


Fig. 10: (a): Ω versus V at $T = 0$, variation of U , (b): Ω versus V at $U = 5.2$, variation with T , (c): the resulting phase diagram.

Figure 10a then shows $\Omega(V)$ at $T = 0$ for different values of U . For smaller U there are two stationary points: a maximum at $V = 0$ and a minimum at finite V , which is the physical solution. At $U_c \approx 5.85$ the two extrema coalesce into a single minimum at $V = 0$, which is the only stationary point for larger U . This change from finite V to $V = 0$ precisely corresponds to the metal-insulator transition. To see this we note that in the special case of $T = 0$ and $\mu = U/2$ the self-energy of the dimer can be evaluated exactly [17]:

$$\Sigma(\omega) = \frac{U}{2} + \frac{U^2}{8} \left(\frac{1}{\omega + 3V} + \frac{1}{\omega - 3V} \right). \quad (30)$$

Note that this has exactly the form (11) derived by Luttinger [12]. The k integrated Green's function then is (note that $\mu = U/2$)

$$G(\omega) = \int_{-2}^2 d\epsilon \frac{\rho_0(\epsilon)}{\omega + U/2 - \epsilon - \Sigma(\omega)}. \quad (31)$$

For real ω the single particle spectral density, i.e. the combined photoemission and inverse photoemission spectrum, is given by

$$A(\omega) = -\frac{1}{\pi} \lim_{\delta \rightarrow 0} \Im G(\omega + i\delta).$$

This is shown in Figure 11 for different V , together with the imaginary part of the self-energy. Since we only want to see qualitatively the effect of vanishing V , $U = 5$ was kept throughout. Then $\Im \Sigma(\omega)$ shows two Lorentzian peaks located at $\pm 3V$ as expected from (30). Each of these peaks creates a gap in the density of states, so that there are three regions with nonvanishing spectral density. As $V \rightarrow 0$ the two poles of $\Sigma(\omega)$ approach each other and the spectral weight in the inner region around $\omega = 0$ which corresponds to the Fermi energy, becomes smaller and smaller. Eventually, at $V = 0$ the two peaks merge and there is no more spectral weight

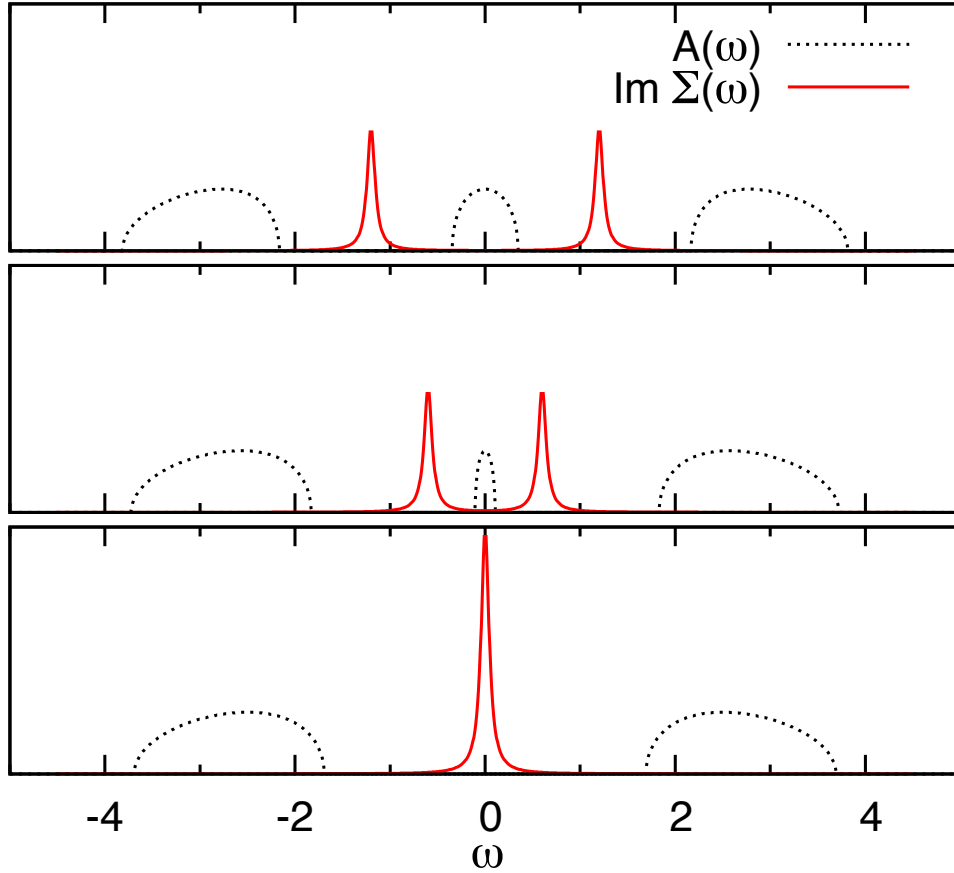


Fig. 11: Single particle spectral function and imaginary part of the self-energy (calculated with an imaginary part of 0.05 for the frequency) obtained from the angle integrated Green's function (31) and the self-energy (30). Parameter values are $U = 5$ and $V = 0.4$ (top), $V = 0.2$ (center) and $V = 0$ (bottom).

at the chemical potential; the system is an insulator now, which is the scenario predicted by dynamical mean-field theory (DMFT) [18]. It remains to be mentioned that DMFT calculations find $U_c \approx 5.84$ [19].

Next, Figure 10b shows $\Omega(V)$ for the fixed value of $U = 5.2$ and different temperatures T . For most temperatures there are three stationary points whereby the local maximum can be discarded. It follows that there are actually two possible solutions for each temperature between $T = 0.10$ and $T = 0.12$. This implies that there is a 1st order phase transition between these two temperatures. Repeating the procedure for various U gives the phase diagram in Figure 10c. There is only a metallic solution for small U , at a first U_{c1} a second insulating solution starts to appear, at U_c there is a first order metal-insulator transition and on from U_{c2} there is only an insulating solution. The results obtained in this way by the, essentially analytical, solution of a dimer are qualitatively very similar to those obtained by extensive numerical renormalization group [20] and quantum Monte Carlo [21] calculations in the framework of DMFT. The main deficiency of the dimer calculation is the underestimation of the critical temperature T_c in Figure 10c by about a factor of two.

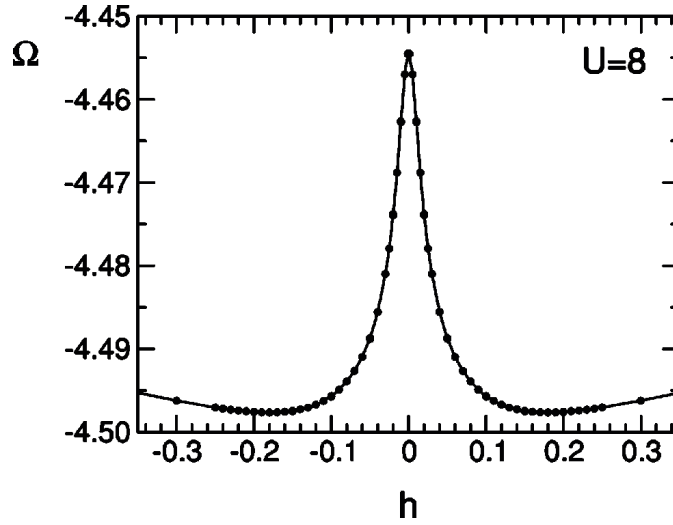


Fig. 12: Variation of Ω with h in (32) for the half-filled 2D Hubbard model. The reference system is an array of 10-site clusters. Reprinted with permission from [22], Copyright 2004 by the American Physical Society.

5.2 Discussion of spontaneous symmetry breaking

As already mentioned, the VCA gives an estimate of the grand potential Ω . This property makes the VCA of particular usefulness for the discussion of ordering transitions. For definiteness, let us assume we want to discuss antiferromagnetism in the 2D Hubbard model and let us assume that we partition the planar model into an array of finite clusters as in Figure 8. Then, since the single-particle terms of the reference system are completely arbitrary, we may include a term

$$\tilde{H}_S = h \sum_j e^{i\mathbf{Q} \cdot \mathbf{R}_j} (n_{j,\uparrow} - n_{j,\downarrow}) \quad (32)$$

with $\mathbf{Q} = (\pi, \pi)$ into \tilde{H}_0 . This term represents a staggered magnetic field which breaks the spin-rotation symmetry of the Hamiltonian. It has to be stressed, however, that *no* magnetic field whatsoever is added to the Hamiltonian of the *lattice system*. The self-energy $\check{\Sigma}(\omega)$ computed in the reference system with $h \neq 0$, however, incorporates this broken symmetry in various ways; for example the self-energy for the two spin-directions will be different and the sites of the cluster are divided into inequivalent sublattices. If we now determine the optimum value of the parameter h from the standard requirement

$$\frac{\partial \Omega_{latt}}{\partial h} = 0$$

there are two possible outcomes: we will usually always find a solution with $h^* = 0$, which corresponds to the paramagnetic state. It may happen, however, that there is a second solution with $h^* \neq 0$, see Figure 12 for an example, and if this gives a lower Ω it follows that even in the complete absence of any magnetic field the grand potential of the lattice system can be lowered by a self-energy which incorporates broken symmetry. The lattice system thus undergoes a transition to a state of spontaneously broken symmetry. In this way, not only various kinds of magnetic ordering but also superconductivity, charge, or orbital ordering can be discussed, see the review by Potthoff [6].

5.3 Photoemission spectra of NiO, CoO and MnO

Lastly, we consider beyond-band-structure calculations for realistic models of $3d$ transition metal compounds. Very often these have a rock-salt structure, such as NiO, CoO and MnO, or a Perovskite structure, such as LaCoO_3 . In both cases the transition metal ion is surrounded by an octahedron of oxygen ions.

It is well-known that these materials are often not well described by LDA calculations and it is widely accepted that the reason is the strong Coulomb-repulsion between electrons in the $3d$ shells of the transition metal ions. The description of this Coulomb interaction is the subject of *multiplet theory*, which was initiated in the 1920's to explain the optical spectra of atoms and ions in the gas phase. Multiplet theory is discussed in many textbooks of atomic physics, in particular the books by Slater [23] and Griffith [24] should be mentioned, as well as [25].

We assume that the orbitals which describe the $3d$ electrons in the Hamiltonian (2) are analogous to atomic wave functions in that they can be labeled by the set of quantum numbers $\nu = (n, l, m, \sigma)$ where $n = 3$ is the principal quantum number, $l = 2$ the total orbital angular momentum quantum number, $m \in \{-l, \dots, l\}$ the z -component of orbital angular momentum, and $\sigma = \pm 1/2$ the z -component of spin. n and l could be omitted because they are identical for all $3d$ orbitals, but we keep them to stay consistent with Slater and Griffith. We introduce creation and annihilation operators $d_{i,\nu}^\dagger$ and $d_{i,\nu}$ for electrons in the $3d$ shell of the transition metal ion i . The Coulomb interaction between the $3d$ electrons then can be written as

$$\begin{aligned}
 H_1 &= \frac{1}{2} \sum_i \sum_{\nu_1, \nu_2, \nu_3, \nu_4} V(\nu_1, \nu_2, \nu_3, \nu_4) d_{i,\nu_1}^\dagger d_{i,\nu_2}^\dagger d_{i,\nu_3} d_{i,\nu_4} \\
 V(\nu_1, \nu_2, \nu_3, \nu_4) &= \delta_{\sigma_1, \sigma_4} \delta_{\sigma_2, \sigma_3} \delta_{m_1+m_2, m_3+m_4} \\
 &\quad \times \sum_{k=0}^{\infty} c^k(l_1 m_1; l_4 m_4) c^k(l_3 m_3; l_2 m_2) R^k(n_1 l_1, n_2 l_2, n_3 l_3, n_4 l_4).
 \end{aligned} \tag{33}$$

Here the *Gaunt coefficients* $c^k(lm; l'm')$ are given by

$$c^k(lm; l'm') = \sqrt{\frac{4\pi}{2k+1}} \int_0^{2\pi} d\phi \int_{-1}^1 d\cos(\Theta) Y_{lm}^*(\Theta, \phi) Y_{k, m-m'}(\Theta, \phi) Y_{l', m'}(\Theta, \phi) \tag{34}$$

and the *Slater integrals* R^k by

$$R^k(n_1 l_1, n_2 l_2, n_3 l_3, n_4 l_4) = e^2 \int_0^\infty dr r^2 \int_0^\infty dr' r'^2 R_{n_1 l_1}(r) R_{n_2 l_2}(r') \frac{r_{<}^k}{r_{>}^{k+1}} R_{n_4 l_4}(r) R_{n_3 l_3}(r'). \tag{35}$$

The Gaunt coefficients are pure numbers, which do not depend on the specific ion and are tabulated in textbooks [23–25]. The calculation of the Slater integrals requires knowledge of the radial wave function $R_{3,2}(r)$ of the $3d$ shell, which is often obtained from Hartree-Fock wave functions for the free transition metal ion in question. In any case, the parameters c^k and R^k can be assumed to be known. More detailed analysis shows, moreover, that for a d -shell only the terms with $k = 0, 2, 4$ in the sum in (33) differ from zero; the sum thus is finite and the Coulomb matrix elements can be calculated without problems. The noninteracting part of the

Hamiltonian can be written as

$$\begin{aligned}
 H_0 = & \sum_{i,j} \sum_{\nu,\lambda} \left(t_{(i,\nu),(j,\lambda)} d_{i,\nu}^\dagger p_{j,\lambda} + H.c. \right) + \sum_i \sum_{\nu_1,\nu_2} C_{\nu_1,\nu_2} d_{i,\nu_1}^\dagger d_{i,\nu_2} \\
 & + \sum_{i,j} \sum_{\lambda_1,\lambda_2} t_{(i,\lambda_1),(j,\lambda_2)} p_{i,\lambda_1}^\dagger p_{j,\lambda_2} .
 \end{aligned} \tag{36}$$

The first term describes hybridization between the $3d$ -orbitals and orbitals on other atoms, which are created by $p_{j,\lambda}^\dagger$ where λ is shorthand for some set of quantum numbers which specify these orbitals. The second term contains the orbital energies of the d -electrons and the effects of the crystalline electric field. The third term describes hybridization between orbitals other than the $3d$ orbitals. The matrix elements $t_{(i,\nu),(j,\lambda)}$ and $t_{(i,\lambda_1),(j,\lambda_2)}$ can be expressed in terms of relatively few parameters such as $(pd\sigma)$, $(pd\pi)$. . . by use of the Slater-Coster tables [26]. For a given compound the parameters in (36) can be obtained, e.g., by a fit to an LDA band structure. It was shown in the pioneering work by Fujimori and Minami [27] that the *momentum integrated* photoemission spectra of transition metal oxides can be reproduced very well by considering an octahedron-shaped cluster comprising only a single transition metal ion and its six nearest neighbor oxygen ions. If only the transition metal $3d$ and the oxygen $2p$ shells are taken into account such a cluster has $5 + 6 \cdot 3 = 23$ orbitals per spin direction. This number can be reduced considerably by noting that in octahedral symmetry for each of the five $3d$ orbitals there is precisely one linear combination of O- $2p$ orbitals on the neighboring oxygen atoms that hybridizes with it, so that the number of relevant orbitals is only 10 per spin direction, which is well manageable by the Lanczos algorithm. The Hamiltonian for the cluster reads

$$H = \sum_{\alpha,\sigma} (t_\alpha d_{\alpha,\sigma}^\dagger p_{\alpha,\sigma} + H.c.) + \sum_{\alpha,\beta,\sigma} c_{\alpha,\beta} d_{\alpha,\sigma}^\dagger d_{\beta,\sigma} + \sum_{\alpha,\beta,\sigma} \tilde{c}_{\alpha,\beta} p_{\alpha,\sigma}^\dagger p_{\beta,\sigma} + H_1 . \tag{37}$$

Here $p_{\alpha,\sigma}^\dagger$ create electrons in the bonding combinations of O- $2p$ orbitals and H_1 is given in (33). The finding of Fujimori and Minami immediately suggests an obvious generalization of Potthoff's treatment of the single-band Hubbard model: instead of a dimer consisting of a single correlated site and a single bath site. see Figure 9 and the Hamiltonian (29), we use an octahedron-shaped cluster comprising the 5 correlated $3d$ orbitals and 5 bath sites corresponding to the the bonding combinations of oxygen $2p$ orbitals, i.e., precisely the Hamiltonian (37) as reference system. The larger size of the clusters makes the calculation more demanding in that the eigenstates of the reference system and the Green's functions now have to be obtained by the Lanczos algorithm. Moreover, the reference system contains more than just one parameter so that (28) actually represents a system of coupled nonlinear equations. The problem still is manageable, however, for the necessary numerical procedures and possible algorithms for the solution of (28) see Refs. [28] and [29].

Here we proceed to some of the results. Figure 13 shows angle-integrated valence band photoemission spectra for the three transition metal oxides NiO, CoO, and MnO. For each compound the figure compares the computed spectral density with transition metal $3d$ character and for oxygen $2p$ character to experimental valence band photoemission spectra obtained with high

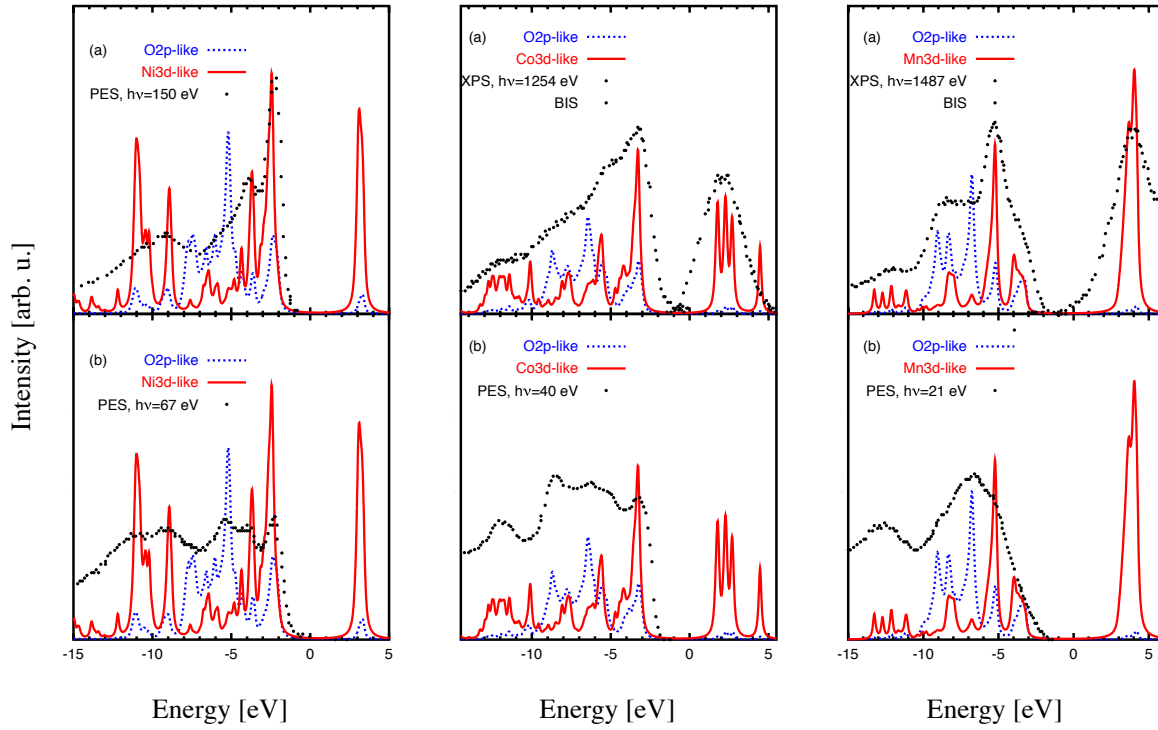


Fig. 13: Angle integrated valence-band photoemission spectra obtained by the VCA for transition metal oxides NiO (left), CoO (center), and MnO (right), compared to experimental spectra taken with high and low photon energy. Reprinted with permission from [28], Copyright 2008 by the American Physical Society.

(top) and relatively low photon energy (bottom). It can be seen that the experimental spectra change substantially with photon energy, and the main reason is the dependence of the photoionization cross section on photon energy [30]. As a rule of thumb, one may say that at X-ray energies the spectra show predominantly the transition-metal 3d-like spectral density, whereas it is the oxygen 2p-like spectral density at low photon energy. Taking this into account, there is good overall agreement between the theoretical and experimental spectra. One may also compare k-resolved spectra and also find good agreement [28].

6 Summary

In summary, Potthoff's new idea of introducing a reference system to generate trial self-energies [4] allows one to combine the classic field theoretical work of Luttinger and Ward [1] with the numerical technique of exact diagonalization of finite systems, resulting in a method for treating strongly correlated lattice systems by exact diagonalization: the variational cluster approximation. Its variational nature makes the VCA particularly useful as exemplified by the 'dimer-DMFT' description of the metal-insulator transition. Since the VCA always gives an estimate for the grand potential, it is particularly useful for treating ordering transitions. By combining this with the very successful cluster method for transition metal oxides [27], it allows one to perform electronic structure calculations using realistic models of transition metal oxides.

7 Appendix: A theorem on determinants

Here we prove the identity

$$\frac{\partial \ln(\det A)}{\partial A_{ij}} = A_{ji}^{-1}.$$

We use Laplace's formula and expand $\det(A)$ in terms of minors

$$\det(A) = \sum_{l=1,n} (-1)^{i+l} A_{il} M_{il}.$$

Since none of the minors M_{il} contains the element A_{ij} we find

$$\frac{\partial \ln(\det A)}{\partial A_{ij}} = \frac{(-1)^{i+j} M_{ij}}{\det(A)}.$$

Next, the i^{th} column of A^{-1} is the solution of the system of equations

$$Ac = e_i,$$

where e_i is the i^{th} column of the unit matrix, which has all elements equal to zero, except for the i^{th} , which is one. We use Kramer's rule and find for the j^{th} element of the i^{th} column

$$A_{ji}^{-1} = \frac{\det(\bar{A}_j)}{\det(A)},$$

where \bar{A}_j is the matrix where the j^{th} column has been replaced by e_i . Now we use again Laplace's formula for $\det(\bar{A}_j)$ and obtain

$$A_{ji}^{-1} = \frac{(-1)^{i+j} M_{ij}}{\det(A)},$$

which proves the theorem.

As an application we assume that the matrix elements of A are functions of some parameter α .

We then find

$$\frac{\partial \ln(\det A)}{\partial \alpha} = \sum_{i,j} \frac{\partial \ln(\det A)}{\partial A_{ij}} \frac{\partial A_{ij}}{\partial \alpha} = \sum_{i,j} A_{ji}^{-1} \frac{\partial A_{ij}}{\partial \alpha} = \text{Tr} \left(A^{-1} \frac{\partial A}{\partial \alpha} \right).$$

References

- [1] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960)
- [2] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)
- [3] L. Hedin, Phys. Rev. **139**, A796 (1965)
- [4] M. Potthoff, Eur. Phys. J. B **32**, 429 (2003)
- [5] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003)
- [6] M. Potthoff: *Self-Energy-Functional Theory*
in A. Avella and F. Mancini (eds.): *Strongly Correlated Systems* (Springer (2012); see also preprint arXiv:11082183
- [7] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinski:
Methods of Quantum Field Theory in Statistical Physics
(Prentice-Hall, New Jersey 1964)
- [8] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems*
(McGraw-Hill, San Francisco, 1971)
- [9] J.W. Negele and H. Orland: *Quantum Many-Particle Systems*
(Addison-Wesley, Redwood, 1988)
- [10] The main difference as compared to FW is that due to the LCAO-like formulation we are using discrete momentum sums rather than integrals i.e. as compared to FW one has to replace $\frac{1}{(2\pi)^3} \int d\mathbf{k} \rightarrow \frac{1}{N} \sum_{\mathbf{k}}$. Similarly, the volume of the crystal is simply $V = N$.
- [11] G. Baym and N. D. Mermin, J. Math. Phys. **2**, 232 (1961)
- [12] J.M. Luttinger, Phys. Rev. **121**, 942 (1961)
- [13] M. Potthoff, Condens. Mat. Phys. **9**, 557 (2006)
- [14] E. Koch: *The Lanczos Method*
in E. Pavarini, E. Koch, A. Lichtenstein, and D. Vollhardt (eds.):
The LDA+DMFT approach to strongly correlated materials
Modeling and Simulation Vol. 1 (Forschungszentrum Jülich, 2011)
<http://www.cond-mat.de/events/correl11>
- [15] M. Aichhorn, E. Arrigoni, M. Potthoff, and W. Hanke Phys. Rev. B **74**, 024508 (2006)
- [16] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003)
- [17] E. Lange, Mod. Phys. Lett. B **12**, 915 (1998)

- [18] A. Georges, G. Kotliar, W. Krauth and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [19] G. Moeller, Q. Si, G. Kotliar, M. Rozenberg, and D.S. Fisher, *Phys. Rev. Lett.* **74**, 2082 (1995)
- [20] R. Bulla, T.A. Costi, and D. Vollhardt, *Phys. Rev. B* **64** 045103 (2001)
- [21] J. Joo and V. Oudovenko, *Phys. Rev. B* **64** 103102 (2001)
- [22] C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni, and M. Potthoff, *Phys. Rev. B* **70**, 245110 (2004)
- [23] J.C. Slater: *Quantum Theory of Atomic Structure* (McGraw-Hill, New York, 1960)
- [24] J.S. Griffith: *The Theory of Transition Metal Ions* (Cambridge University Press, 1961)
- [25] R. Eder: *Multiplets in Transition Metal Ions*
in E. Pavarini, E. Koch, F. Anders, and M. Jarrell:
Correlated Electrons: From Models to Materials
Modeling and Simulation Vol. 2 (Forschungszentrum Jülich, 2012)
<http://www.cond-mat.de/events/correl12>
- [26] W.A. Harrison: *Electronic Structure and the Properties of Solids* (Dover, 1989)
- [27] A. Fujimori and F. Minami, *Phys. Rev. B* **30**, 957 (1984)
- [28] R. Eder, *Phys. Rev. B* **78**, 115111 (2008)
- [29] R. Eder, *Phys. Rev. B* **81**, 035101 (2010)
- [30] D.E. Eastman and J.L. Freeouf, *Phys. Rev. Lett.* **34**, 395 (1975)

5 Magnetism: From Stoner to Hubbard

Alexander Lichtenstein

I. Institut für Theoretische Physik

Universität Hamburg, 20355 Hamburg, Germany

Contents

1	Introduction	2
2	From Stoner to Hubbard	2
3	From LDA to DMFT	3
4	Realistic DMFT scheme	10
5	Solution of quantum impurity problem	12
5.1	Hirsch-Fye quantum Monte Carlo	12
5.2	Continuous-time quantum Monte Carlo	14
5.3	Fluctuation exchange approximation	18
6	Effective magnetic interactions in LDA+DMFT	19
7	LDA+DMFT results for itinerant ferromagnetic metals	21
8	Conclusions	30

1 Introduction

We will discuss a realistic approach to magnetism and electronic structure of correlated materials which takes into account dynamical many-body effects. The scheme combines the features of the itinerant electron theory (Stoner) of magnetic crystals with the localized-moment description (Heisenberg) in a unified spin-fluctuations approach for a generalized multi-orbital Hubbard model. Moreover, we analyze the calculation of effective exchange interaction-parameters based on the realistic electronic structure of correlated magnetic crystals.

2 From Stoner to Hubbard

We start to discuss the different models of magnetic materials (Fig. 1) with the simplest one-band Stoner Hamiltonian

$$H_s = \sum_{\mathbf{k}\sigma} (\varepsilon_{\mathbf{k}} + I\langle n_{-\sigma} \rangle) c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma}, \quad (1)$$

where $\varepsilon_{\mathbf{k}}$ is the energy band spectrum and I is a Stoner interatomic exchange parameter. In this case the temperature-dependent magnetic properties are related to the so-called Stoner excitations from the occupied "spin-up" to the unoccupied "spin-down" band. They reduce the magnitude of the magnetization, so that finally at the Curie point the itinerant system becomes a nonmagnetic metal.

If we compare the Stoner model with a standard Hubbard approach with the Hamiltonian

$$H_h = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_i U n_{i\uparrow} n_{i\downarrow}, \quad (2)$$

where t_{ij} are the hopping parameters and U the characteristic Hubbard Coulomb interaction, then one can easily realize that the Stoner model is just a mean-field approximation to the Hubbard model. In the weakly correlated case the only possible magnetic excitations are spin-flips, and the corresponding energy is of the order $I \cdot M$ with $M = \langle n_\uparrow - n_\downarrow \rangle$ which is much larger than realistic Curie temperatures. In the opposite limit, the strongly correlated Hubbard model at half-filling [1], one can derive an effective Heisenberg model

$$H_e = - \sum_{ij} J_{ij} \vec{S}_i \cdot \vec{S}_j. \quad (3)$$

The kinetic exchange interactions $J_{ij} = -2t_{ij}t_{ji}/U$ are of the order of magnetic (Néel) transition temperatures. The Heisenberg model describes well the magnetism of localized $4f$ -materials. In the case of transition metals, where both longitudinal and transverse magnetic fluctuations are important, the most appropriate model is the Hubbard Hamiltonian, Eq. (2).

We can discuss the different approaches to estimate the effective Heisenberg interactions, presented in Figs. 2 to 4. In Fig. 2 a simple two-site spin-model for the Heisenberg interaction with the singlet and triplet states is compared with the so-called Slater one-electron model for antiferromagnetic states, which results in an additional factor of two in the definition of the effective

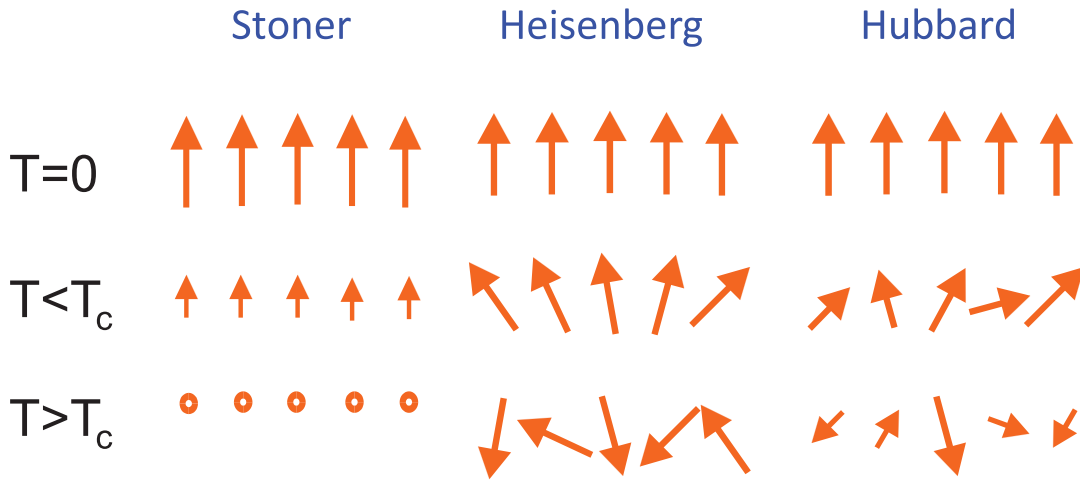


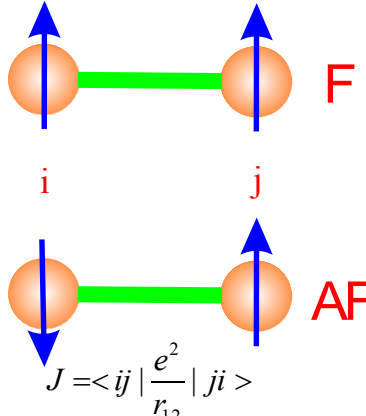
Fig. 1: Schematic view of different models of magnetism: Stoner model for itinerant weakly correlated electrons, Heisenberg model for localized magnetic moments and Hubbard model for the spin-fluctuations model of correlated electrons.

exchange interaction. In Fig. 3 the solution of the two-site Hubbard model for the many-body sector with one spin-up and one spin-down electron is shown, which results in the famous Anderson kinetic exchange interaction [1]. Finally, one can show that the mean-field solution of the Hubbard model with band energies modified via infinitesimal spin-rotations results in the same effective exchange interactions of a classical Heisenberg model (Fig. 4). In this case we used the so-called “local force theorem” which was originally formulated for density functional theory [2] and will be proven later (Sec. 6) using the Baym-Kadanoff approach. This theorem gives a simple recipe to obtain the total energy difference for a small perturbation of the charge- or spin-density as a change from the non-selfconsistent band energy for corresponding perturbation. In our case the energy of the infinitesimal spin-rotation in the two-site mean-field rotationally invariant Slater (spin-polarized LDA) model have been calculated and compared with the corresponding classical Heisenberg model (Fig. 4). We can see that both schemes give exactly the same effective exchange parameter.

3 From LDA to DMFT

The calculation of thermodynamic properties and excitation spectra of different magnetic materials is one of the most important problems of the microscopic theory of magnetism. We introduce a general functional approach which will cover density-functional theory (DFT), dynamical mean-field theory (DMFT) and Baym-Kadanoff (BK) theories [3]. Let us start from the full many-body Hamiltonian describing the electrons moving in the periodic external potential $V(\mathbf{r})$ of the ions with chemical potential μ and interacting via the Coulomb law $U(\mathbf{r}-\mathbf{r}') = 1/|\mathbf{r}-\mathbf{r}'|$. We use atomic units $\hbar = m = e = 1$. In the field-operator representation the Hamiltonian has

Magnetism of H₂: Heisenberg vs. Slater



Exchange interaction:

$$H_{ex} = -2J_{ij} \vec{S}_i \vec{S}_j$$

$$J = \frac{1}{2}(E_S - E_T) = E_{AF} - E_F$$

$$E_F = |\uparrow\uparrow\rangle = E_T$$

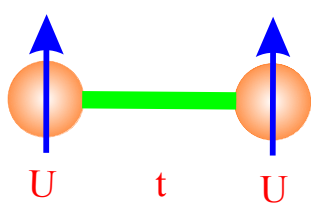
$$E_{AF} = |\uparrow\downarrow\rangle = \frac{1}{2}(E_S + E_T)$$

$$\psi_T = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle); \quad |\uparrow\uparrow\rangle; \quad |\downarrow\downarrow\rangle$$

$$\psi_S = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$$

Fig. 2: Exchange interaction in a two-site Heisenberg model.

Exchange interactions: Anderson



$$H = t \sum_{ij=1,2} c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_{i=1,2} n_{i\uparrow} n_{i\downarrow}$$

ED in subspace
 $N_{\uparrow}=1, N_{\downarrow}=1$

N	i	j	H	1	2	3	4
1	↑	↓	1	0	0	t	t
2	↓	↑	2	0	0	t	t
3	↑↓	0	3	t	t	U	0
4	0	↑↓	4	t	t	0	U

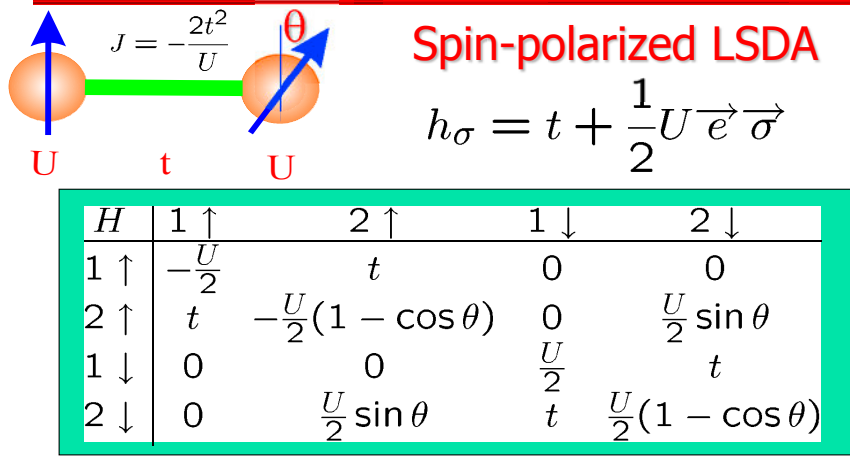
$U \gg t$

Anderson kinetic exchange

$$J = -\frac{2t^2}{U}$$

Fig. 3: Exchange interaction in the two-site Hubbard model.

Exchange: Local force approach



Spectrum: $\epsilon_i = \pm \frac{1}{2} \sqrt{4t^2 + U^2 \pm Ut \sqrt{2(1 \mp \cos \theta)}}$

Exchange energy: $E_x(\theta) \approx \frac{t^2}{U} \cos \theta = -2JS^2 \vec{e}_i \cdot \vec{e}_j = -\frac{1}{2}J \cos \theta$

Fig. 4: Exchange interaction in the two-site mean-field rotationally invariant Slater model.

the form

$$H = \sum_{\sigma} \int d\mathbf{r} \hat{\psi}_{\sigma}^{\dagger}(\mathbf{r}) \left[-\frac{1}{2} \nabla^2 + V(\mathbf{r}) - \mu \right] \hat{\psi}_{\sigma}(\mathbf{r}) + \frac{1}{2} \sum_{\sigma\sigma'} \int d\mathbf{r} \int d\mathbf{r}' \hat{\psi}_{\sigma}^{\dagger}(\mathbf{r}) \hat{\psi}_{\sigma'}^{\dagger}(\mathbf{r}') U(\mathbf{r} - \mathbf{r}') \hat{\psi}_{\sigma'}(\mathbf{r}') \hat{\psi}_{\sigma}(\mathbf{r}). \quad (4)$$

We can always use a single-particle orthonormal basis set $\phi_n(\mathbf{r})$, in solids for example Wannier orbitals with a full set of quantum numbers, e.g., site, orbital and spin index $n = (i, m, \sigma)$, and expand the fields in creation and annihilation operators

$$\begin{aligned} \hat{\psi}(\mathbf{r}) &= \sum_n \phi_n(\mathbf{r}) \hat{c}_n \\ \hat{\psi}^{\dagger}(\mathbf{r}) &= \sum_n \phi_n^*(\mathbf{r}) \hat{c}_n^{\dagger} \end{aligned} \quad (5)$$

Going from fermionic operators to the Grassmann variables $\{c_n^*, c_n\}$ we can write the functional integral representation for the partition function of the many-body Hamiltonian in the imaginary time domain using the Euclidean action S

$$Z = \int \mathcal{D}[c^*, c] e^{-S} \quad (6)$$

$$S = \sum_{12} c_1^* (\partial_{\tau} + t_{12}) c_2 + \frac{1}{2} \sum_{1234} c_1^* c_2^* U_{1234} c_4 c_3, \quad (7)$$

where the one- and two-electron matrix elements are defined as

$$\begin{aligned} t_{12} &= \int d\mathbf{r} \phi_1^*(\mathbf{r}) \left[-\frac{1}{2} \nabla^2 + V(\mathbf{r}) - \mu \right] \phi_2(\mathbf{r}) \\ U_{1234} &= \int d\mathbf{r} \int d\mathbf{r}' \phi_1^*(\mathbf{r}) \phi_2^*(\mathbf{r}') U(\mathbf{r} - \mathbf{r}') \phi_3(\mathbf{r}) \phi_4(\mathbf{r}'). \end{aligned} \quad (8)$$

and we use the following short definition of the sum: $\sum_1 \dots \equiv \sum_{im} \int d\tau \dots$

The one-electron Green function is defined via the simplest non-zero correlation function for fermions

$$G_{12} = -\langle c_1 c_2^* \rangle_S = -\frac{1}{Z} \int \mathcal{D}[c^*, c] c_1 c_2^* \exp(-S) \quad (9)$$

and gives all information on the spin-dependent electronic structure of correlated materials.

The main difficulties of strongly interacting electronic systems are related to the fact that the higher-order correlation functions do not separate into a product of lower-order correlation functions. For example the two-particle Green function or generalized susceptibility (X) is defined in the following way [4]

$$X_{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_S = \frac{1}{Z} \int \mathcal{D}[c^*, c] c_1 c_2 c_3^* c_4^* \exp(-S), \quad (10)$$

and can be expressed as a simple "non-interacting" part and a connected correlated contribution through the exact Green function and the full vertex function Γ_{1234} [5]

$$X_{1234} = G_{14} G_{23} - G_{13} G_{24} + \sum_{1'2'3'4'} G_{11'} G_{22'} \Gamma_{1'2'3'4'} G_{3'3} G_{4'4}. \quad (11)$$

In principle, the spin-dependent part of the two-particle correlation function or generalized magnetic susceptibility contains all information on the magnetic properties of solids.

Modern computational material science is based on the density-functional (DFT) approach [6]. It is a common practice to use this scheme not only for the total energy calculations and related quantities such as charge and spin densities, but also for different spectral characteristics. Sometimes the agreement of the computational results with the experimental data is very impressive, despite the absence of a reliable theoretical background. In principle, the energies of Kohn-Sham quasiparticles [6] which are calculated in standard band theory are just auxiliary quantities for the total energy calculation.

The DFT functional is defined in the following way: the Kohn-Sham potential $V_{KS} = V_{ext} + V_H + V_{xc}$ plays the role of the effective one-electron potential, where all exchange-correlations effects V_{xc} are taken into account. In this case, V_{ext} is the external potential and V_H is the Hartree potential. In principle the exchange-correlation potential V_{xc} is known only for the homogeneous electron gas [8]. Therefore in practical applications one uses the so-called local-density approximation (LDA) to DFT. The total-energy functional reads

$$E_{tot}[n] = T_0[n] + V_{ext}[n] + V_H[n] + V_{xc}[n] \quad (12)$$

where T_0 is the kinetic energy of the non-interacting systems. Finally, if we define the total electron density as

$$n(\mathbf{r}) = \sum_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) \quad (13)$$

the local-density approximation to the DFT reads

$$T_0[n] + V_{ext}[n] = \sum_i \int d\mathbf{r} \phi_i^*(\mathbf{r}) \left[-\frac{1}{2} \nabla^2 + V_{ext}(\mathbf{r}) - \mu \right] \phi_i(\mathbf{r}) \quad (14)$$

$$E_H[n] = \frac{1}{2} \int d\mathbf{r} n(\mathbf{r}) U(\mathbf{r} - \mathbf{r}') n(\mathbf{r}') \quad (15)$$

$$E_{xc}[n] = \int d\mathbf{r} n(\mathbf{r}) \varepsilon_{xc}(n(\mathbf{r})) \quad (16)$$

where $\varepsilon(n)$ is exchange correlation density for the homogeneous electron gas, which has been calculated with quantum Monte Carlo (QMC) [8]. The variational principle leads to the Kohn-Sham effective one-electron equation (in atomic units)

$$\left[-\frac{1}{2} \nabla^2 + V_{ext}(\mathbf{r}) + V_H + V_{xc} \right] \phi_k(\mathbf{r}) = \varepsilon_k \phi_k(\mathbf{r}). \quad (17)$$

In the one-band case the back Fourier transform of $\varepsilon_{\mathbf{k}}$ will give an effective Kohn-Sham hopping parameter $t(\mathbf{R})$ to t_{ij} , where i, j are the lattice indices. For the realistic multi-orbital case one can use the efficient first-principle Wannier-function parameterization of the energy bands to get the multi orbital (m) hopping matrix elements $t_{im,jm'}^{\text{LDA}}$, which will be used in the magnetic many-body formalism.

In the DFT scheme we lose information about the non-equal time Green function, which gives the single particle excitation spectrum as well as the \mathbf{k} -dependence of the spectral function, and restrict ourselves only to the ground state energy of the many-electron system. Moreover, we also lose information about all collective excitations in solids, such as plasmons or magnons, which can be obtained from the generalized susceptibility.

Despite all achievements of the quantitative electronic structure theory, the list of difficulties and shortcomings is growing, especially when considering the magnetic d - and f -electron systems. In a number of cases the theory appears even *qualitatively* inadequate. First, the DFT scheme cannot describe correctly the phenomenon of “Mott insulators” [7], as was first observed by Terakura *et al.* [9] in their attempt of calculating the electronic structure of $3d$ -metal oxides. Later we faced similar problems in field of high- T_c superconductors [10] and other compounds [11]. The Ce- and U- based “heavy fermion” compounds such as CeCu_6 , UPt_3 , etc, are other “hot-spots”: normally the calculated effective masses are orders of magnitude smaller than what is experimentally observed [12]. Even for the pure $3d$ -metals some qualitative differences between theory and experiment exists. For example, there are at least three difficulties with the photoelectron spectra of ferromagnetic nickel [13]: (i) the measured width of the occupied part of the d -band is 30% narrower than calculated (ii) the spin-splitting is half of what is predicted by LDA and (iii) the band structure cannot describe the famous 6 eV satellite. Calculations for

paramagnetic spin-disordered states [14] lead to the conclusion that Ni has no local magnetic moments above the Curie temperature T_C , in clear contradiction with experimental results [15]. For iron, standard band theory cannot explain the data about spin polarization of thermionic electrons [16–18] and some features of angle-resolved photoemission spectra [18–20]. All these difficulties raise questions about the DFT approach: what is the “electron spectrum” that we really calculate and how can we improve the electronic theory for magnetic d - and f -systems? It was understood many years ago that all these problems are connected with the inadequate description of many-body effects in DFT calculations of the excitation spectra. Methods such as GW [21] and simplified $LDA+U$ [11] have been proposed to improve the situation. These methods are very useful for the description of antiferromagnetic transition-metal oxides as Mott insulators [11]. However, one should note that $LDA+U$ is just a mean-field approximation and cannot describe *correlation* effects which are, by definition, the many-body effects *beyond* Hartree-Fock. For example, in these approaches one needs spin- or orbital-ordering to describe the Mott insulator and it is impossible to describe correctly the electronic structure of NiO or MnO in the paramagnetic phase. At the same time, the magnetic ordering should not be important for the basic physics of Mott insulators [7]. All the “Hartree-Fock-like” approaches fail to describe the renormalization of the effective mass in the heavy fermion systems. There are also many problems concerning the electronic structure and itinerant magnetism of $3d$ metals as described above. Thus, one needs some practical ways of incorporating *correlation* effects in the electronic structure of solids.

In principle, there are two ways to include them into DFT calculations. The first uses a *time-dependent* DFT formalism which can guarantee, in principle, an opportunity to calculate exact response functions [22], in the same sense as the Hohenberg-Kohn theorem guarantees the total energy in usual “static” DFT [6]. However, all the expressions for this time-dependent non-local DFT in real calculations are based on RPA-like approximations which do not give a satisfactory description of really highly correlated systems. They are excellent for investigating the plasmon spectrum of aluminum, but not for understanding the nature of high- T_c superconductivity or heavy fermion behavior. Another way is to use an “alternative” many-body theory developed in the 50s by Gell-Mann and Brueckner, Galitskii and Migdal, Beliaev and many others in terms of the Green functions rather than the electron density [23]. We try to formulate such a computational approach as a generalization of $LDA+U$ scheme, the so-called “ $LDA+DMFT$ ” method. The main difference between $LDA+DMFT$ and $LDA+U$ is that in the former dynamical fluctuations, the real correlation effects, are accounted for by a local but energy dependent self-energy $\Sigma(\omega)$.

A comparison of the standard DFT theory in the local-density approximation (LDA) and the $LDA+DMFT$ approach is represented in table I. First of all, LDA is based on the Hohenberg-Kohn theorem stating that the total energy E_{tot} is a functional of charge (and spin) densities, while the $LDA+DMFT$ scheme considers the thermodynamic potential Ω as a functional of exact one-particle Green functions. This approach in many-particle theory has been introduced in the works by Luttinger and Ward [24] and Baym and Kadanoff [25]. The Green function in $LDA+DMFT$ plays the same role as the density matrix in LDA. We stress the dynamical

Table 1: *Comparison of LDA and realistic DMFT schemes*

LDA	LDA+DMFT
Density functional	Baym-Kadanoff functional
Density $\rho(\mathbf{r})$	Green-Function $G(\mathbf{r}, \mathbf{r}', \omega)$
Potential $V_{xc}(\mathbf{r})$	Self-energy $\Sigma_i(\omega)$
$E_{tot} = E_{sp} - E_{dc}$	$\Omega = \Omega_{sp} - \Omega_{dc}$
$E_{sp} = \sum_{k < k_F} \varepsilon_k$	$\Omega_{sp} = -\text{Tr} \ln[-G^{-1}]$
$E_{dc} = E_H + \int \rho V_{xc} d\mathbf{r} - E_{xc}$	$\Omega_{dc} = \text{Tr} \Sigma G - \Phi_{LW}$

nature of the correlation effects that are taken into account in the LDA+DMFT approach since the density in the LDA is just the static limit of the local Green function. Further, the self-energy Σ is treated analogously to the exchange-correlation potential, the local approximation for Σ , which is assumed to be energy-dependent but not momentum-dependent, corresponds to the local approximation for V_{xc} . In both formalisms the thermodynamic potential can be represented in a “single-particle” form, Ω_{sp} minus the contributions of the “doubly counted terms”, Ω_{dc} . This will be important for the consideration of the so-called “local force theorem” and the computation of magnetic interaction parameters. The single-particle contribution to the thermodynamic potential in the LDA+DMFT would have the same form as in the LDA if we were taking into account only the poles of the Green function and neglected the quasiparticle damping. However, even then the quasiparticle energies are not quite the same since the poles of the Green functions do not coincide, generally speaking, with the “Kohn-Sham” energies. The quantity Φ_{LW} is the Luttinger-Ward generating functional for the self energy, or the sum of all the skeleton diagrams without free legs [24].

The difficulty with finite temperature effects is one of the main shortcomings of a standard DFT formalism. In all realistic calculations the temperature is included in the Fermi distribution functions and in the lattice constants via the thermal expansion [26]. At the same time, for the itinerant electron magnets the temperature effects connected with the “Bose” degrees of freedom due to spin waves and paramagnons are much more important [27]. In principle, these effects could be taken into account in DFT via the temperature dependence of the exchange-correlation potential, the corresponding terms being nonlocal. It is not easy to propose an adequate expression for such temperature-dependent non-local potential. One of the first attempts in this direction is based on simple RPA-like considerations [28]. On the other hand, in LDA+DMFT-type scheme all calculations are naturally carried out for finite temperatures by using Matsubara frequencies, as is usual in many-body theory [23].

The main assumption of the LDA+DMFT approach is the importance of only intra-site “Hubbard correlations” with the local approximation for the self-energy. It is worthwhile to stress a

difference of this kind of locality from the locality in DFT theory. In the latter, the local approximation means that the exchange-correlation energy is calculated for the homogeneous electron gas [8]. It is known from exact QMC calculations that the correlation effects lead to some instabilities of the state of homogeneous electron gas (magnetism, charge ordering, etc) only for electron densities which are orders of magnitude smaller than ones typical for real metals (the critical values of the parameter r_s are of order of hundred in comparison with the “normal” range 2-6 for metals). At the same time, magnetism and charge ordering are rather usual for real compounds with the d - and f -elements. It thus seems that the “atomic-like” features of d - and f -states are of crucial importance to describe the correlation effects in real compounds. Only these features are taken into account in the Hubbard-like terms for the d - or f -states in LDA+DMFT approach. Therefore one can view the LDA+DMFT as the simplest way for quantitative considerations of correlation effects in transition metals and their compounds, based on the LDA description for all non-correlated electrons in the systems.

The investigation of correlation effects in the electronic structure and magnetism of iron-group metals is still far from having found the final picture and attracts continuous interest (see, e.g., [14, 29–31] and references therein). Despite many attempts, the situation is still unclear both theoretically and experimentally. For example, there is no agreement on the presence of a 5 eV satellite in the photoemission spectrum of iron [19, 20], and on the existence of local spin splitting above the Curie temperature in nickel [32]. From the theoretical point of view, different approaches such as second-order perturbation theory [30, 33], the T -matrix approximation [29, 34], the three-body Faddeev approximation [35], and the moment expansion method [36] were used. Unfortunately, the range applicability of these schemes is not clear. Here we present the LDA+DMFT approach [18, 37, 38] which is based on the combination of standard band-theory techniques (LDA) with dynamical mean-field theory (DMFT) [39].

4 Realistic DMFT scheme

In the LDA+DMFT approach we consider the renormalisation of the “bare” LDA energy or electron hopping due to correlation effects. Of course the t_{ij} contain already some part of the correlation effects but only those which may be considered in the local density approximation. The most important “rest” in strongly correlated system is the correlations of the Hubbard type [41] due to the intra-site Coulomb repulsion. Therefore we start from the general form of the LDA+DMFT Hamiltonian

$$H = \sum_{ij\sigma\{m\}} t_{im_1,jm_2}^{\text{LDA}} c_{im_1\sigma}^\dagger c_{jm_2\sigma} + \frac{1}{2} \sum_{i\{\sigma m\}} U_{m_1 m_2 m'_1 m'_2}^i c_{im_1\sigma}^\dagger c_{im_2\sigma'}^\dagger c_{im'_2\sigma'} c_{im'_1\sigma} \quad (18)$$

where the (i, j) represents different crystal sites, $\{m\}$ label different orbitals and the $\{\sigma\}$ are spin indices. Coulomb matrix elements are defined in the usual way (see Eq. (8) with the screened Coulomb interactions in the basis of localized Wannier functions).

The simplified form of the LDA+DMFT Hamiltonian is related to the diagonal “density-density

approximation”

$$H = \sum_{\{im\sigma\}} t_{im,i'm'}^{\text{LDA}} c_{im\sigma}^\dagger c_{i'm'\sigma} + \frac{1}{2} \sum_{imm'\sigma} U_{mm'}^i n_{im\sigma} n_{im'-\sigma} + \frac{1}{2} \sum_{im \neq m'\sigma} (U_{mm'}^i - J_{mm'}^i) n_{im\sigma} n_{im'\sigma} \quad (19)$$

where i is the site index and m the orbital quantum numbers, $\sigma = \uparrow, \downarrow$ the spin projection, c^\dagger, c the Fermi creation and annihilation operators ($n = c^\dagger c$), and t^{LDA} is the effective single-particle Hamiltonian obtained from the non-magnetic LDA with the correction for double counting of the average interactions among d -electrons. In the general case of a spin-polarized LDA Hamiltonian this correction is presented in Refs. [18, 37, 40]. In the magnetic LDA it is just a shift “back” of correlated d -states with respect to the s and p -states by the average Coulomb and exchange potential

$$E_{dc} = \frac{1}{2} \bar{U} n_d (1 - n_d) - \frac{1}{2} \bar{J} [n_{d\uparrow} (1 - n_{d\uparrow}) + n_{d\downarrow} (1 - n_{d\downarrow})]$$

with \bar{U} and \bar{J} being the average Coulomb and exchange interactions and $n_d = n_{d\uparrow} + n_{d\downarrow}$ the total number of correlated $d(f)$ -electrons.

The screened Coulomb and exchange vertices for the d -electrons

$$U_{mm'} = \langle mm' | U_{scr}(\mathbf{r} - \mathbf{r}') | mm' \rangle \quad (20)$$

$$J_{mm'} = \langle mm' | U_{scr}(\mathbf{r} - \mathbf{r}') | m'm \rangle \quad (21)$$

are expressed via the effective Slater integrals. We use the minimal spd -basis in the LMTO-TB formalism [42] and numerical orthogonalization for the $t^{\text{LDA}}(\mathbf{k})$ matrix [37]. The local density approximation [6] is used for the self-consistent electronic structure calculation.

In order to find the best local approximation for the self-energy we use the DMFT method [43] for real systems. This scheme becomes exact in the limit of infinite lattice coordination number [44]. The DMFT approach reduces the lattice many-body problem (Eq. (21)) to the self-consistent solution of an effective one-site Anderson model. In this case we need a local Green-function matrix which has the following form in the orthogonal Wannier representation

$$G(i\omega) = \sum_{\mathbf{k}} \{ (i\omega + \mu) \hat{1} - t^{\text{LDA}}(\mathbf{k}) - \Sigma(i\omega) \}^{-1}, \quad (22)$$

where μ is the chemical potential. Note that due to the cubic crystal symmetry of ferromagnetic bcc -iron the local Green function in the absence of spin-orbit interaction is diagonal both in the orbital and the spin indices. The so-called bath Green function that defines the effective Anderson model and preserves the double-counting of the local self-energy is obtained as a solution of the impurity model via [43]

$$\mathcal{G}^{-1}(i\omega) = G^{-1}(i\omega) + \Sigma(i\omega) \quad (23)$$

5 Solution of quantum impurity problem

5.1 Hirsch-Fye quantum Monte Carlo

As discussed above, DMFT maps the many-body system onto a multi-orbital quantum impurity, i.e., a set of local degrees of freedom in a bath described by the Weiss field function \mathcal{G} . The impurity action (here $n_{m\sigma} = c_{m\sigma}^\dagger c_{m\sigma}$ and $\mathbf{c}(\tau) = [c_{m\sigma}(\tau)]$ is a vector of Grassman variables) is given by:

$$S_{imp} = - \int_0^\beta d\tau \int_0^\beta d\tau' \text{Tr}[\mathbf{c}^\dagger(\tau) \mathcal{G}^{-1}(\tau, \tau') \mathbf{c}(\tau')] + \frac{1}{2} \sum_{m, m', \sigma} \int_0^\beta d\tau [U_{mm'} n_\sigma^m n_{-\sigma}^{m'} + (U_{mm'} - J_{mm'}) n_\sigma^m n_\sigma^{m'}]. \quad (24)$$

It describes the spin-, orbital-, energy-, and temperature-dependent interactions of a particular magnetic 3d-atom with the rest of the crystal and is used to compute the local Green function matrix

$$\mathbf{G}_\sigma(\tau - \tau') = -\frac{1}{Z} \int D[\mathbf{c}, \mathbf{c}^\dagger] e^{-S_{imp}} \mathbf{c}(\tau) \mathbf{c}^\dagger(\tau'), \quad (25)$$

where Z is the partition function.

The local Green functions for the imaginary time interval $[0, \beta]$ with the mesh $\tau_l = l\Delta\tau$, $l = 0, \dots, L-1$, and $\Delta\tau = \beta/L$, where $\beta = 1/T$ are calculated in the path-integral formalism [43,45]

$$G_m^{ll'} = \frac{1}{Z} \sum_{s_{mm'}^l} \det[O(s)] * G_m^{ll'}(s). \quad (26)$$

Here we redefined for simplicity $m \equiv \{m, \sigma\}$, and the so-called fermion-determinant $\det[O(s)]$ as well as the Green function for an arbitrary set of auxiliary fields $G(s) = O^{-1}(s)$ are obtained via the Dyson equation [46] for the imaginary-time matrix ($\mathbf{G}_m(s) \equiv G_m^{ll'}(s)$):

$$\mathbf{G}_m = [\mathbf{1} - (\mathbf{G}_m^0 - \mathbf{1})(e^{V_m} - \mathbf{1})]^{-1} \mathbf{G}_m^0,$$

where the effective fluctuation potential from the Ising fields $s_{mm'}^l = \pm 1$ is

$$V_m^l = \sum_{m'(\neq m)} \lambda_{mm'} s_{mm'}^l \sigma_{mm'}, \text{ where } \sigma_{mm'} = \begin{cases} 1, m < m' \\ -1, m > m' \end{cases}$$

and the discrete Hubbard-Stratonovich parameters are $\lambda_{mm'} = \text{arccosh}[\exp(\frac{1}{2}\Delta\tau U_{mm'})]$ [46]. Using the output local Green function from QMC and input bath Green functions the new self-energy is obtain via Eq. (23) and the self-consistent loop can be closed through Eq. (22). The main problem of the multiband QMC formalism is the large number of the auxiliary fields $s_{mm'}^l$. For each time slice l it is equal to $M(2M-1)$ where M is the total number of orbitals, giving 45 Ising fields for a d -shell. We computed the sum over these auxiliary fields in Eq.(26) using importance-sampling QMC, and performed a dozen self-consistent iterations over the

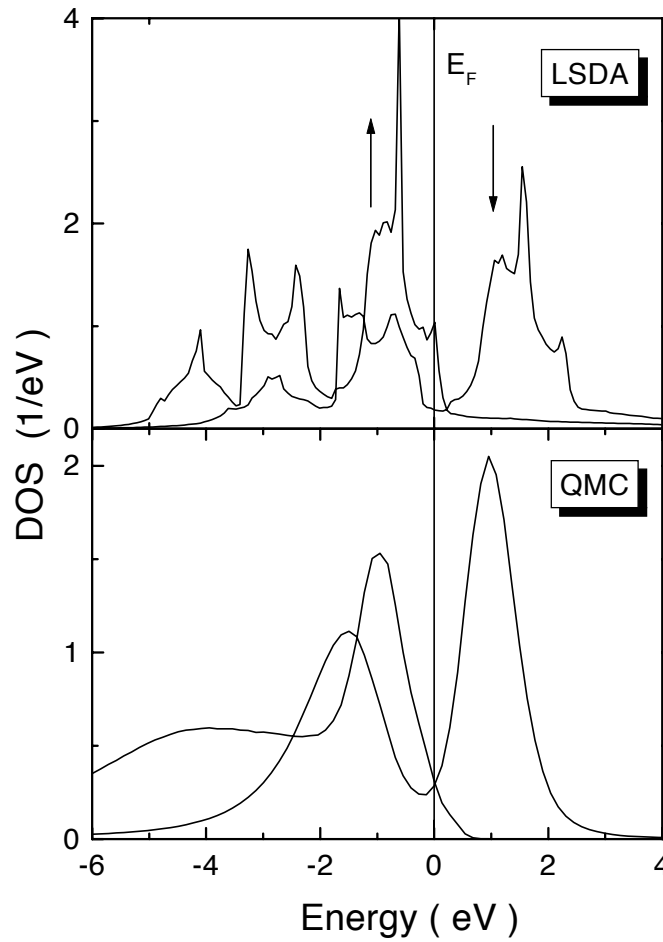


Fig. 5: Spin-resolved density of d -states and magnetic moments for ferromagnetic iron calculated in LDA (top) and LDA+QMC (bottom) for different average Coulomb interactions with $J=0.9$ eV and temperature $T=850$ K.

self-energy Eqs. (22,23,26). The number of QMC sweeps was of the order of 10^5 on the CRAY-T3e supercomputer. The final $G_m(\tau)$ has very little statistical noise. We use the maximum entropy method [47] for analytical continuations of the QMC Green functions to the real axis.

Comparison of the total density of states (DOS) with the results of LDA calculations (Fig. 5) shows reasonable agreement for the single-particle properties of the not “highly correlated” ferromagnetic iron. We calculate bcc -iron at its experimental lattice constant with 256 k -points in the irreducible part of Brillouin zone. The Matsubara summation corresponds to the temperature of about $T = 850$ K. The average magnetic moment is about $1.9 \mu_B$, which corresponds to a small reduction of the LDA-value of $2.2 \mu_B$ for such a high temperature. The DOS curves in the LDA+ Σ approach with the exact QMC solution of on-site multi-orbital problem is similar to that obtained within the simple perturbative fluctuation-exchange (FLEX) approximation described below. The discussion of the results and the comparison with experimental data will be given in Section 4.

5.2 Continuous-time quantum Monte Carlo

Even though DMFT reduces the extended lattice problem to a single-site problem, the solution of the underlying Anderson impurity model remains a formidable quantum many-body problem that requires accurate solvers. Recently a new class has emerged, the continuous-time quantum impurity solvers. These are based on stochastic, i.e., Monte-Carlo methods and mainly come in two different flavors: The weak and strong-coupling approach.

The weak-coupling or interaction expansion (CT-INT) continuous-time quantum Monte Carlo algorithm for fermions was originally introduced by Aleksey Rubtsov [48]. The power of the new CT-QMC scheme is that it performs the path integral without any transformation to effective non-interacting models and that it can be used for any complicated electron-electron vertex. We introduce the algorithm in the path integral formulation for the single-orbital Anderson impurity problem with a Hubbard-type interaction $Un_{\uparrow}n_{\downarrow}$. The generalization to the multiorbital case can be found in Ref. [49]. To start, the action for the Anderson impurity model, up to an irrelevant additive constant, is divided into a Gaussian S_0 and an interacting part S_U

$$S_0 = \sum_{\sigma} \int_0^{\beta} d\tau \int_0^{\beta} d\tau' c_{\sigma}^*(\tau) [\partial_{\tau} - \mu + \Delta(\tau - \tau') + U\alpha_{-\sigma}(\tau)\delta(\tau - \tau')] c_{\sigma}(\tau'), \quad (27)$$

$$S_U = U \int_0^{\beta} d\tau [c_{\uparrow}^*(\tau)c_{\uparrow}(\tau) - \alpha_{\uparrow}(\tau)][c_{\downarrow}^*(\tau)c_{\downarrow}(\tau) - \alpha_{\downarrow}(\tau)]. \quad (28)$$

The parameters α are introduced to control the sign problem. A formal series expansion for the partition function is obtained by expanding the exponential in the interaction term,

$$\mathcal{Z} = \int e^{-S_0[c^*, c]} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} U^k \int_0^{\beta} d\tau_1 \dots \int_0^{\beta} d\tau_k [c_{\uparrow}^*(\tau_1)c_{\uparrow}(\tau_1) - \alpha_{\uparrow}(\tau_1)][c_{\downarrow}^*(\tau_1)c_{\downarrow}(\tau_1) - \alpha_{\downarrow}(\tau_1)] \dots [c_{\uparrow}^*(\tau_k)c_{\uparrow}(\tau_k) - \alpha_{\uparrow}(\tau_k)][c_{\downarrow}^*(\tau_k)c_{\downarrow}(\tau_k) - \alpha_{\downarrow}(\tau_k)] \mathcal{D}[c^*, c]. \quad (29)$$

Using the definition of the average over the noninteracting action:

$$\langle \dots \rangle_0 = \frac{1}{\mathcal{Z}_0} \int \mathcal{D}[c^*, c] \dots \exp(-S_0), \quad (30)$$

the partition function can be expressed in the following form

$$\mathcal{Z} = \mathcal{Z}_0 \sum_{k=0}^{\infty} \int_0^{\beta} d\tau_1 \dots \int_{\tau_{k-1}}^{\beta} d\tau_k \text{sgn}(\Omega_k) |\Omega_k|, \quad (31)$$

where the integrand is given by

$$\Omega_k = (-1)^k U^k \langle [c_{\uparrow}^*(\tau_1)c_{\uparrow}(\tau_1) - \alpha_{\uparrow}(\tau_1)][c_{\downarrow}^*(\tau_1)c_{\downarrow}(\tau_1) - \alpha_{\downarrow}(\tau_1)] \dots [c_{\uparrow}^*(\tau_k)c_{\uparrow}(\tau_k) - \alpha_{\uparrow}(\tau_k)][c_{\downarrow}^*(\tau_k)c_{\downarrow}(\tau_k) - \alpha_{\downarrow}(\tau_k)] \rangle_0. \quad (32)$$

Note that here the range of time integration has been changed such that time ordering is explicit: $\tau_1 < \dots < \tau_{k-1} < \tau_k$. For a given set of times all $k!$ permutations of this sequence

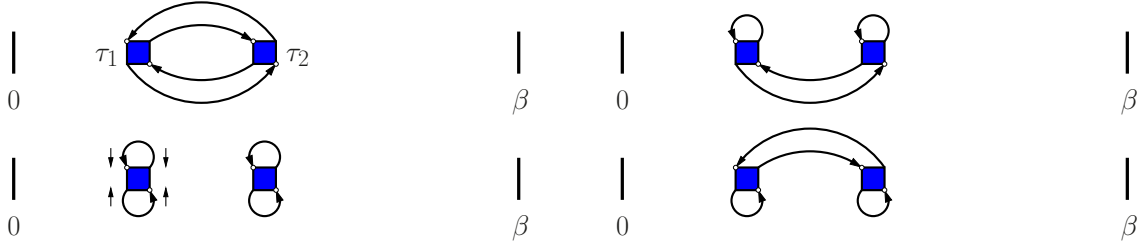


Fig. 6: The four contributions to the partition function for $k = 2$. The interaction vertices are depicted by squares. Bare Green functions are shown as lines. Vertical arrows indicate the spin direction. Connecting the vertices by Green functions in all possible ways is the interpretation of the determinant.

contribute to Eq. 29. These can be brought into the standard sequence by permuting quadruples of Grassmann numbers, and hence without gaining an additional sign. Since all terms are subject to time-ordering, their contribution to the integral is identical, so that the factor $1/k!$ in (29) cancels. A configuration can be fully characterized by specifying a perturbation order k and an (unnumbered) set of k times: $C_k = \{\tau_1, \dots, \tau_k\}$.

The Monte Carlo algorithm performs importance sampling over this configuration space. The weight of a configuration is thereby taken to be equal to the modulus of the integrand, Eq. 32. Since S_0 is Gaussian, the average over the noninteracting system can be evaluated using Wick's theorem. Hence the weight of a configuration is essentially given by a fermionic determinant of a matrix containing the bare Green functions

$$\Omega_k = (-1)^k U^k \prod_{\sigma} \det \hat{g}^{\sigma}, \quad (33)$$

where the local Green function in the α fields is equal to

$$(\hat{g}^{\sigma})_{ij} = g_0^{\sigma}(\tau_i - \tau_j) - \alpha_{\sigma}(\tau_i) \delta_{ij}. \quad (34)$$

Note that determinants for different spin orientations factorize since the Green function is diagonal in spin-space.

The hybridization expansion (CT-HYB) or strong-coupling algorithm was initially introduced by Philipp Werner *et al.* [50] and has been generalized to multi-orbital systems with general interactions [51, 52]. Here the algorithm is discussed in the segment representation, which allows for a very fast computation of the trace for density-density type interactions. The action is regrouped into the atomic part

$$S_{\text{at}} = \int_0^{\beta} d\tau \sum_{\sigma} c_{\sigma}^*(\tau) [\partial_{\tau} - \mu] c_{\sigma}(\tau) + U \int_0^{\beta} d\tau c_{\uparrow}^*(\tau) c_{\uparrow}(\tau) c_{\downarrow}^*(\tau) c_{\downarrow}(\tau) \quad (35)$$

and the part of the action S_{Δ} which contains the hybridization term:

$$S_{\Delta} = - \int_0^{\beta} d\tau' \int_0^{\beta} d\tau \sum_{\sigma} c_{\sigma}(\tau) \Delta(\tau - \tau') c_{\sigma}^*(\tau'). \quad (36)$$

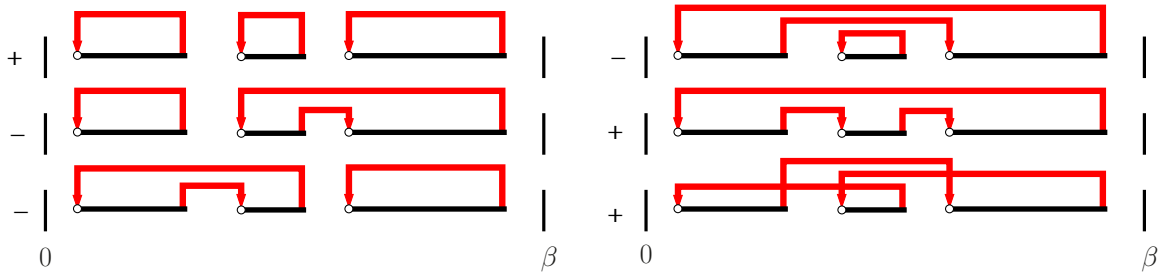


Fig. 7: Diagrammatic representation of the six contributions to the partition function for spinless fermions at $k = 3$. An electron is inserted at the start of a segment (marked by an open circle) and removed at the segment endpoint. The hybridization function lines $\Delta(\tau_i - \tau'_j)$ (shown in red) are connected to the segments in all possible ways. The sign of each diagram is given on the left. The diagrams can be collected into a determinant. Reproduced from Ref. [50].

Here the sign is taken out by reversing the original order of c and c^* to avoid an alternating sign in the expansion. To simplify the notation, consider first the spinless fermion model, which is obtained by disregarding the spin sums and interaction in Eqs. (35), (36). The series expansion for the partition function is generated by expanding in the hybridization term

$$\mathcal{Z} = \int \mathcal{D}[c^*, c] e^{-S_{\text{at}}} \sum_k \frac{1}{k!} \int_0^\beta d\tau'_1 \int_0^\beta d\tau_1 \dots \int_0^\beta d\tau'_k \int_0^\beta d\tau_k \times \\ \times c(\tau_k) c^*(\tau'_k) \dots c(\tau_1) c^*(\tau'_1) \Delta(\tau_1 - \tau'_1) \dots \Delta(\tau_k - \tau'_k). \quad (37)$$

The important observation now is that, at any order, the diagrams can be collected into a determinant of hybridization functions. The partition function then takes the form

$$\mathcal{Z} = \mathcal{Z}_{\text{at}} \sum_k \int_0^\beta d\tau'_1 \int_{\tau'_1}^\beta d\tau_1 \dots \int_{\tau_{k-1}}^\beta d\tau'_k \int_{\tau'_k}^{\circ\tau'_k} d\tau_k \langle c(\tau_k) c^*(\tau'_k) \dots c(\tau_1) c^*(\tau'_1) \rangle_{\text{at}} \det \hat{\Delta}^{(k)}, \quad (38)$$

where the average is over the states of the atomic problem described by S_{at} . Here $\det \hat{\Delta}^{(k)}$ denotes the determinant of the matrix of hybridizations $\hat{\Delta}_{ij} = \Delta(\tau_i - \tau'_j)$. The diagrams contributing to the partition function for $k = 3$ are shown in Fig. 7. A diagram is depicted by a collection of segments, where a segment is symbolic for the time interval where the impurity is occupied. The collection of diagrams obtained by connecting the hybridization lines in all possible ways corresponds to the determinant. Collecting the diagrams into a determinant is essential to alleviate or completely suppress the sign problem. Note that the imaginary time interval in Eq. (38) is viewed as a circle denoted by $\circ\tau'_k$. The trajectories in the path integral are subject to antiperiodic boundary conditions which is accommodated by an additional sign if a segment winds around the circle.

For the single-orbital Anderson impurity model with Hubbard interaction the segment picture still holds and gives a very intuitive picture of the imaginary time dynamics. A configuration is visualized by two separate timelines, one for each spin. The additional sum over spins, $\sum_{\sigma_1 \dots \sigma_k}$,

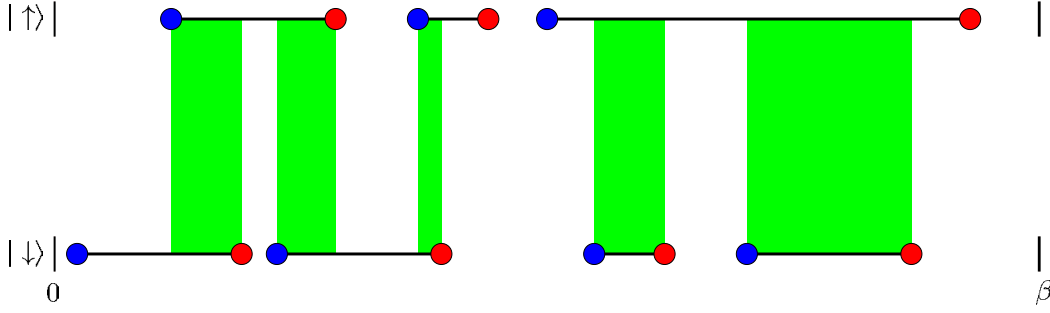


Fig. 8: Example one band CT-HYB: a segment picture: blue dots illustrate annihilation operators, red ones creation operators, and the black line represent the hybridization function $\Delta(\tau_i - \tau_j)$. The green regions represent the time interval at which two electrons are present on the impurity with the total time l_d for which the U price has to be paid.

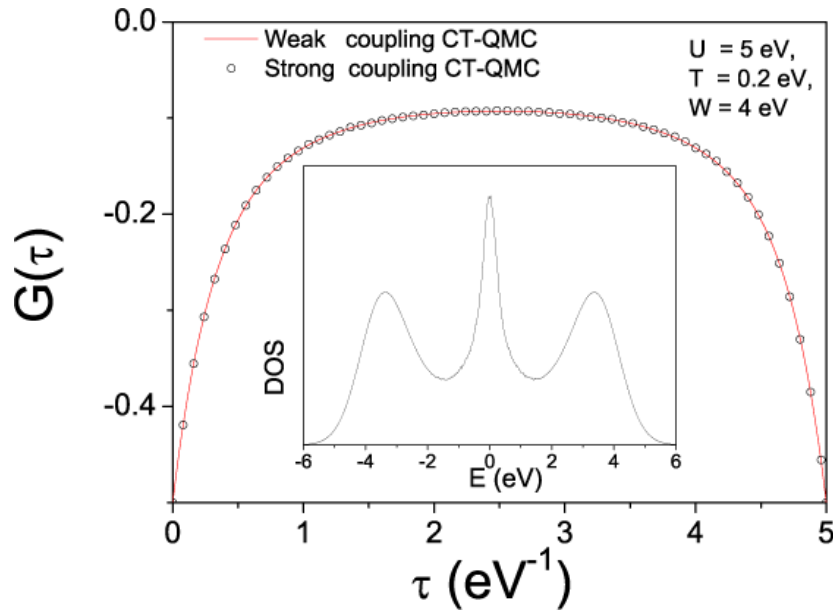


Fig. 9: Comparison of the weak coupling (CT-INT) and strong coupling (CT-HYB) CT-QMC impurity solvers for one-band semicircular model with $U \geq W$. In the insert the density of states obtained with maximum entropy scheme is shown

which enters in the first line of Eq. (38), generates contributions such as the one shown in Fig. 8. The only difference from the spinless fermion model is that when the impurity is doubly occupied, the energy U has to be paid and the trace is $e^{\mu(l_\uparrow + l_\downarrow) - U l_d}$, where l_σ is the time spent on the impurity for an electron with spin σ and l_d is the time the impurity is doubly occupied.

In Fig. 9 we show the comparison of CT-INT and CT-HYB calculations for a strongly coupled ($U \geq W$) single band model. The perfect agreement of these two complementary CT-QMC schemes supports the important conclusion about the possibility of the numerically exact solution of quantum impurity problems.

5.3 Fluctuation exchange approximation

The QMC method described above is probably the most rigorous practical way to solve an effective impurity problem in the framework of DMFT. However, it is rather time consuming. Besides that, in the previous section we did not work with the *complete* four-index Coulomb matrix

$$\langle 12 | v | 34 \rangle = \int d\mathbf{r} d\mathbf{r}' \psi_1^*(\mathbf{r}) \psi_2^*(\mathbf{r}') v_{scr}(\mathbf{r} - \mathbf{r}') \psi_3(\mathbf{r}) \psi_4(\mathbf{r}'), \quad (39)$$

where we define for simplicity $m_1 \equiv 1$.

For moderately strong correlations (which is the case of iron group metals) one can propose an approximate scheme which is more suitable for the calculations. It is based on the fluctuation exchange (FLEX) approximation by Bickers and Scalapino [53], generalized to the multiband spin-polarized case [18, 37, 54]. The electronic self-energy in the FLEX is equal to

$$\Sigma = \Sigma^{\text{HF}} + \Sigma^{(2)} + \Sigma^{(ph)} + \Sigma^{(pp)}, \quad (40)$$

where the Hartree-Fock contribution has a standard form

$$\Sigma_{12,\sigma}^{\text{HF}} = \sum_{34} \left[\langle 13 | v | 24 \rangle \sum_{\sigma'} n_{34}^{\sigma'} - \langle 13 | v | 42 \rangle n_{34}^{\sigma} \right], \quad (41)$$

with the occupation matrix $n_{12}^{\sigma} = G_{21}^{\sigma}(\tau \rightarrow -0)$; this contribution to Σ is equivalent to the spin-polarized “rotationally-invariant” LDA+ U method [40].

The second-order contribution in the spin-polarized case reads

$$\begin{aligned} \Sigma_{12,\sigma}^{(2)}(\tau) = & - \sum_{\{3-8\}} \langle 13 | v | 74 \rangle G_{78}^{\sigma}(\tau) \times \\ & \times \left[\langle 85 | v | 26 \rangle \sum_{\sigma'} G_{63}^{\sigma'}(\tau) G_{45}^{\sigma'}(-\tau) - \langle 85 | v | 62 \rangle G_{63}^{\sigma}(\tau) G_{45}^{\sigma}(-\tau) \right], \end{aligned} \quad (42)$$

and the higher-order particle-hole (or particle-particle) contribution

$$\Sigma_{12,\sigma}^{(ph)}(\tau) = \sum_{34,\sigma'} W_{13,42}^{\sigma\sigma'}(\tau) G_{34}^{\sigma'}(\tau), \quad (43)$$

with $p-h$ ($p-p$) fluctuation potential matrix

$$W^{\sigma\sigma'}(i\omega) = \begin{bmatrix} W^{\uparrow\uparrow}(i\omega) & W^{\uparrow\downarrow}(i\omega) \\ W^{\downarrow\uparrow}(i\omega) & W^{\downarrow\downarrow}(i\omega) \end{bmatrix}, \quad (44)$$

where the spin-dependent effective potentials have a generalized RPA-form and can be found in [18]. Note that for both the $p-h$ and $p-p$ channels the effective interactions, according to Eq. (44), are non-diagonal matrices in *spin space* as in the QMC-scheme, in sharp contrast to any mean-field approximation like LDA. This can be important for spin-dependent transport phenomena in transition metal multilayers.

We could further reduce the computational procedure by neglecting dynamical interactions in the p - p channel since the most important fluctuations in itinerant electron magnets are spin-fluctuations in the p - h channel. We take into account static (of T -matrix type) renormalizations of the effective interactions, replacing the bare matrix $U_{12,34} = \langle 12 | v | 34 \rangle$ in the FLEX-equations with the corresponding spin-dependent scattering T -matrix

$$\langle 12 | T^{\sigma\sigma'} | 34 \rangle = \langle 12 | v | 34 \rangle - \sum_{5678} \langle 12 | v | 56 \rangle \int_0^\beta d\tau G_{56}^\sigma(\tau) G_{78}^{\sigma'}(\tau) \langle 78 | T^{\sigma\sigma'} | 34 \rangle. \quad (45)$$

Similar approximation have been checked for the Hubbard model [55] and appear to be accurate enough for not too large U . Finally, in the spirit of the DMFT-approach $\Sigma = \Sigma[G_0]$, and all the Green functions in the self-consistent FLEX-equations are in fact the bath Green-functions G_0 .

6 Effective magnetic interactions in LDA+DMFT

A useful scheme for analyses of exchange interactions in the LDA approach is the so called “local force theorem”. In this case the calculation of small total energy changes reduces to variations of the one-particle density of states [56, 57]. First of all, let us prove the analog of the local force theorem in the LDA+DMFT approach. In contrast with standard density-functional theory, it deals with the real dynamical quasiparticles defined via Green functions for the correlated electrons rather than with Kohn-Sham “quasiparticles” which are, strictly speaking, only auxiliary states for the total energy calculations. Therefore, instead of working with the thermodynamic potential as a *density* functional we have to start from the general expression for Ω in terms of the exact Green function given in Table I. We have to also keep in mind the Dyson equation $G^{-1} = G_0^{-1} - \Sigma$ and the variational identity $\delta\Phi_{LW} = \text{Tr} \Sigma \delta G$. Here $\text{Tr} = \text{Tr}_{\omega i L \sigma}$ is the sum over Matsubara frequencies $\text{Tr}_\omega \dots = \sum_\omega \dots$, with $\omega = \pi T (2n + 1)$, $n = 0, \pm 1, \dots$, with T the temperature, $iL\sigma$ the numbers of sites (i), the orbital quantum numbers ($L = l, m$), and the spin projections σ , correspondingly. We represent the expression for Ω as a difference of “single particle” (sp) and “double counted” (dc) terms as is usual in density-functional theory. When neglecting the quasiparticle damping, Ω_{sp} will be nothing but the thermodynamic potential of “free” fermions but with exact quasiparticle energies. Suppose we change the external potential, for example, by small spin rotations. Then the variation of the thermodynamic potential can be written as

$$\delta\Omega = \delta^* \Omega_{sp} + \delta_1 \Omega_{sp} - \delta\Omega_{dc} \quad (46)$$

where δ^* is the variation without taking into account the change of the “self-consistent potential”, i.e. self energy, and δ_1 is the variation due to the change of Σ . To avoid a possible misunderstanding, note that we consider the variation of Ω in the general “non-equilibrium” case when the torques acting on spins are nonzero and therefore $\delta\Omega \neq 0$. In order to study the response of the system to general spin rotations one can consider either variations of the spin directions at fixed effective fields or, vice versa, rotations of the effective fields, i.e. variations

of Σ , at fixed magnetic moments. We use the second way. Taking into account the variational property of Φ , it can be easily shown (cf. Ref. [24]) that

$$\delta_1 \Omega_{sp} = \delta \Omega_{dc} = \text{Tr } G \delta \Sigma \quad (47)$$

and hence

$$\delta \Omega = \delta^* \Omega_{sp} = -\delta^* \text{Tr } \ln [\Sigma - G_0^{-1}] \quad (48)$$

which is an analog of the “local force theorem” in density-functional theory [57].

In the LDA+DMFT scheme, the self energy is local, i.e., is diagonal in the site indices. Let us write the spin-matrix structure of the self-energy and Green function in the following form

$$\Sigma_i = \Sigma_i^c + \mathbf{\Sigma}_i^s \cdot \boldsymbol{\sigma}, \quad G_{ij} = G_{ij}^c + \mathbf{G}_{ij}^s \cdot \boldsymbol{\sigma} \quad (49)$$

where $\Sigma_i^{(c,s)} = (\Sigma_i^\uparrow \pm \Sigma_i^\downarrow)/2$ and $\mathbf{\Sigma}_i^s = \Sigma_i^s \mathbf{e}_i$, with \mathbf{e}_i being the unit vector in the direction of effective spin-dependent potential on site i , $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices, $G_{ij}^c = \text{Tr}_\sigma(G_{ij})/2$ and $\mathbf{G}_{ij}^s = \text{Tr}_\sigma(G_{ij}\boldsymbol{\sigma})/2$. We assume that the bare Green function G^0 does not depend on spin directions and all the spin-dependent terms including the Hartree-Fock terms are incorporated into the self-energy. Spin excitations with low energies are connected with the rotations of vectors \mathbf{e}_i

$$\delta \mathbf{e}_i = \delta \boldsymbol{\varphi}_i \times \mathbf{e}_i \quad (50)$$

According to the “local force theorem” (48) the corresponding variation of the thermodynamic potential can be written as

$$\delta \Omega = \delta^* \Omega_{sp} = \mathbf{V}_i \delta \boldsymbol{\varphi}_i, \quad (51)$$

where the torque is equal to

$$\mathbf{V}_i = 2 \text{Tr}_{\omega L} [\mathbf{\Sigma}_i^s \times \mathbf{G}_{ii}^s]. \quad (52)$$

Using the spinor structure of the Dyson equation one can write the Green function in this expression in terms of pair contributions. As a result, we represent the total thermodynamic potential of spin rotations or the effective Hamiltonian in the form [38]

$$\Omega_{spin} = - \sum_{ij} \text{Tr}_{\omega L} \{ (\mathbf{G}_{ij}^s \cdot \mathbf{\Sigma}_j^s) (\mathbf{G}_{ji}^s \cdot \mathbf{\Sigma}_i^s) - \Sigma_i^s G_{ij}^c \Sigma_j^s G_{ji}^c - i (\mathbf{\Sigma}_i^s \times \mathbf{G}_{ij}^c \cdot \mathbf{\Sigma}_j^s) \mathbf{G}_{ji}^s \}. \quad (53)$$

One can show by direct calculation that

$$\left[\frac{\delta \Omega_{spin}}{\delta \boldsymbol{\varphi}_i} \right]_{G=\text{const}} = \mathbf{V}_i. \quad (54)$$

This means that $\Omega_{spin} \{ \mathbf{e}_i \}$ is the effective spin Hamiltonian. The last term in Eq. (53) is nothing but a Dzialoshinskii-Moriya interaction term. It is non-zero only in the relativistic case where $\mathbf{\Sigma}_j^s$ and \mathbf{G}_{ji}^s can be, generally speaking, “non-parallel” and $G_{ij} \neq G_{ji}$ for crystals without inversion center.

In the non-relativistic case, one can rewrite the spin Hamiltonian for small spin deviations near collinear magnetic structures in the following form

$$\Omega_{spin} = - \sum_{ij} J_{ij} \mathbf{e}_i \cdot \mathbf{e}_j, \quad (55)$$

where

$$J_{ij} = -\text{Tr}_{\omega_L} \left(\Sigma_i^s G_{ij}^\uparrow \Sigma_j^s G_{ji}^\downarrow \right) \quad (56)$$

are the effective exchange parameters. This formula generalizes the LDA expressions of [57,58] to the case of correlated systems.

The spin-wave spectrum in ferromagnets can be considered both directly from the exchange parameters or from the energy of corresponding spiral structures (cf. Ref. [57]). In the non-relativistic case when anisotropy is absent one has

$$\omega_{\mathbf{q}} = \frac{4}{M} \sum_j J_{0j} (1 - \cos \mathbf{q} \cdot \mathbf{R}_j) \equiv \frac{4}{M} [J(\mathbf{0}) - J(\mathbf{q})], \quad (57)$$

where M is the magnetic moment (in Bohr magnetons) per magnetic ion.

It should be noted that the expression for the spin-stiffness tensor $D_{\alpha\beta}$ defined by the relation $\omega_{\mathbf{q}} = D_{\alpha\beta} q_\alpha q_\beta$ ($\mathbf{q} \rightarrow 0$) in terms of exchange parameters has to be exact as the consequence of the phenomenological Landau-Lifshitz equations, which are definitely correct in the long-wavelength limit. Direct calculation based on variation of the total energy under spiral spin rotations (cf. Ref. [57]) leads to the following expression

$$D_{\alpha\beta} = -\frac{2}{M} \text{Tr}_{\omega_L} \sum_{\mathbf{k}} \left(\Sigma^s \frac{\partial G^\uparrow(\mathbf{k})}{\partial k_\alpha} \Sigma^s \frac{\partial G^\downarrow(\mathbf{k})}{\partial k_\beta} \right), \quad (58)$$

where \mathbf{k} is the quasi-momentum and the summation is over the Brillouin zone. The expressions Eqs. (56) and (57) are reminiscent of usual RKKY indirect exchange interactions in the s - d exchange model (with Σ^s instead of the s - d exchange integral). One can prove that this expression for the stiffness is exact within the local approximation [59]. At the same time, the exchange parameters themselves, generally speaking, differ from the exact response characteristics defined via static susceptibility since the latter contains vertex corrections. The derivation of approximate exchange parameters from the variations of the thermodynamic potential can be useful for the estimation of J_{ij} in different magnetic systems.

7 LDA+DMFT results for itinerant ferromagnetic metals

We have started from the spin-polarized LDA band structure of ferromagnetic iron within the LMTO method [42] in the minimal s, p, d basis set and used numerical orthogonalization to find the H_t part of our starting Hamiltonian. We take into account Coulomb interactions only between d -states. The correct parameterization of the H_U part is indeed a serious problem. For example, first-principle estimations of average Coulomb interactions (U) [30,60] in iron lead to unreasonably large values of order of 5–6 eV in comparison with experimental values of the U -parameter in the range of 1–2 eV [30]. A semi-empirical analysis of the appropriate interaction value [61] gives $U \simeq 2.3$ eV. The difficulties with choosing the correct value of U are connected with complicated screening problems, definitions of orthogonal orbitals in the crystal, and contributions of the inter-site interactions. In the quasi-atomic (spherical) approximation the full

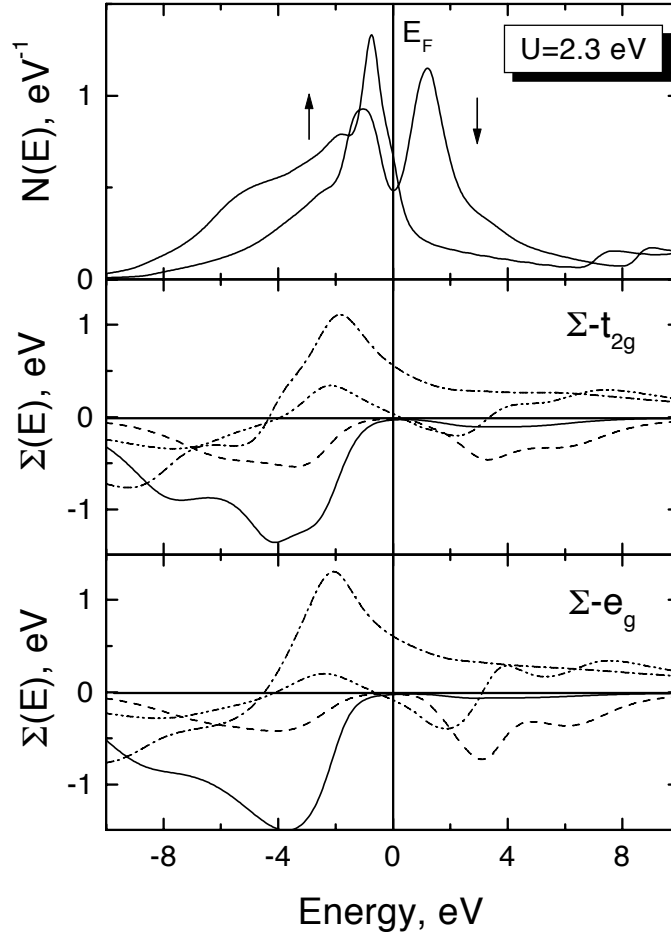


Fig. 10: Total spin-polarized density of states and d -part of the self-energy for iron with $U = 2.3$ eV and $J = 0.9$ eV for the temperature $T = 750$ K. Two different self-energies for t_{2g} and e_g d -states in the cubic crystal field symmetry are presented. The four different lines correspond to imaginary-part spin-up (full line) and spin-down (dashed line) as well as real-part spin-up (dashed-dot line) and spin-down (dashed-double-dot line).

U -matrix for the d -shell is determined by the three parameters U , J and δJ or equivalently by effective Slater integrals F^0 , F^2 and F^4 [11,37]. For example, $U = F^0$, $J = (F^2 + F^4)/14$ and we use the simplest way of estimating δJ , or F^4 , keeping the ratio F^2/F^4 equal to its atomic value 0.625 [62].

Note that the value of the intra-atomic (Hund) exchange interaction J is not sensitive to the screening and approximately equals 0.9 eV in different estimations [60]. For the most important parameter U , which defines the bare vertex matrix Eq. (39), we use the value $U = 2.3$ eV for Fe [61], $U = 3$ eV for Co and Mn and $U = 4$ eV for Ni and Cu. To calculate the spectral functions $A_\sigma(\mathbf{k}, E) = -\text{Tr}_L G_\sigma(\mathbf{k}, E + i0)/\pi$ and DOS as their sum over the Brillouin zone, we first performed analytical continuations for the matrix self-energy from Matsubara frequencies to the real axis using the Pade approximation [63], and then numerically inverted the Green-function matrix as in Eq. (22) for each \mathbf{k} -point. In the self-consistent solution of the FLEX equations we used 1024 Matsubara frequencies and the FFT-scheme with an energy cut-off at 100 eV. The

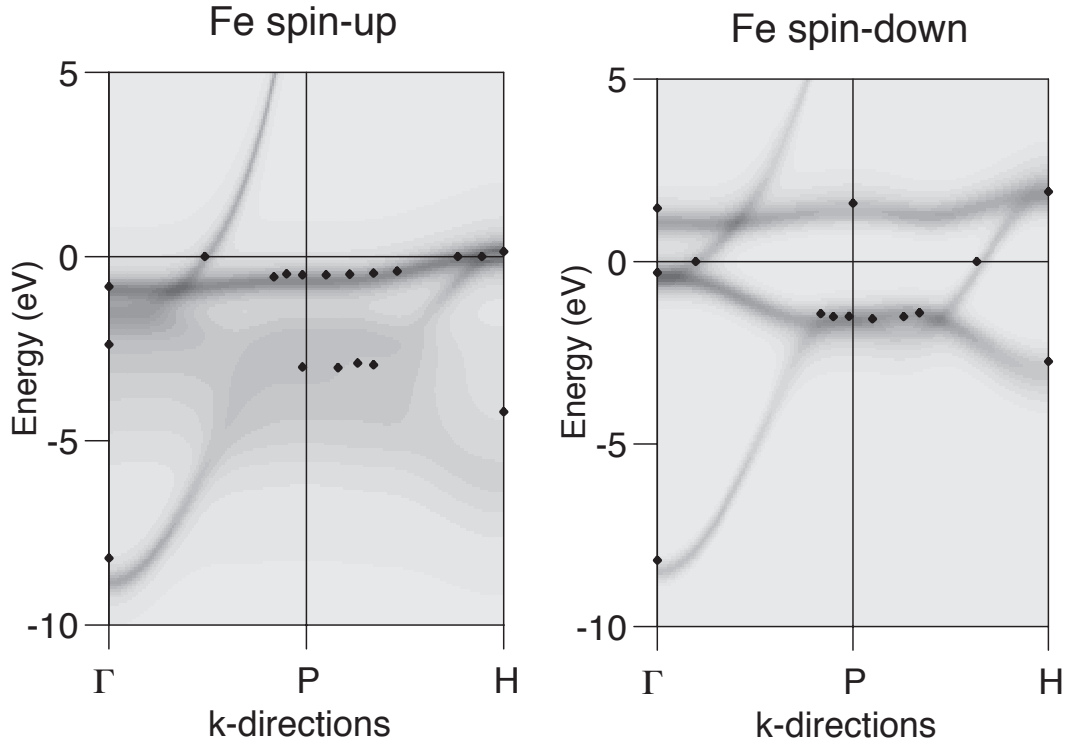


Fig. 11: *Spectral function of ferromagnetic iron for spin-up (a) and spin-down (b) and the two k -directions in the Brillouin zone compared with the experimental angle resolved photoemission and de Haas van Alphen (at the $E_F=0$) points.*

sums over the irreducible Brillouin zone have been made with 72 k -points for SCF-iterations and 1661 k -points for the final total density of states.

The depolarization of states near the Fermi level is another important correlation effect. The decrease of the ratio $P = [N_{\uparrow}(E_F) - N_{\downarrow}(E_F)] / [N_{\uparrow}(E_F) + N_{\downarrow}(E_F)]$ is a typical sign of spin-polaron effects [31, 64]. In our approach these effects are taken into account through the $W_{\uparrow\downarrow}^{(ph)}$ terms in the effective spin-polarized LDA+DMFT potential (Eq. (44)).

The energy dependence of self-energy in Fig. 10 shows the characteristic features of moderately correlated systems. At low energies $|E| < 1$ eV we see a typical Fermi-liquid behavior $\text{Im}\Sigma(E) \sim -E^2$, and $\partial \text{Re}\Sigma(E) / \partial E < 0$. At the same time, for the states beyond this interval within the d -bands the damping is rather large (of the order of 1 eV) so these states correspond to ill-defined quasiparticles, especially for the occupied states. This is probably one of the most important conclusions of our calculations. Qualitatively it was already pointed out in Ref. [33] on the basis of model second-order perturbation-theory calculations. We have shown that this is still the case for the realistic quasiparticle structure of iron with a reasonable value for the Coulomb interaction parameter.

Due to the noticeable broadening of the quasiparticle states, a description of the computational results in terms of effective band structure (determined, for example, from the maximum of the spectral density) would be incomplete. We present in Fig. 11 the *full* spectral density $A_{\sigma}(\mathbf{k}, E)$ including both coherent and incoherent parts as a function of \mathbf{k} and E . We see that in general

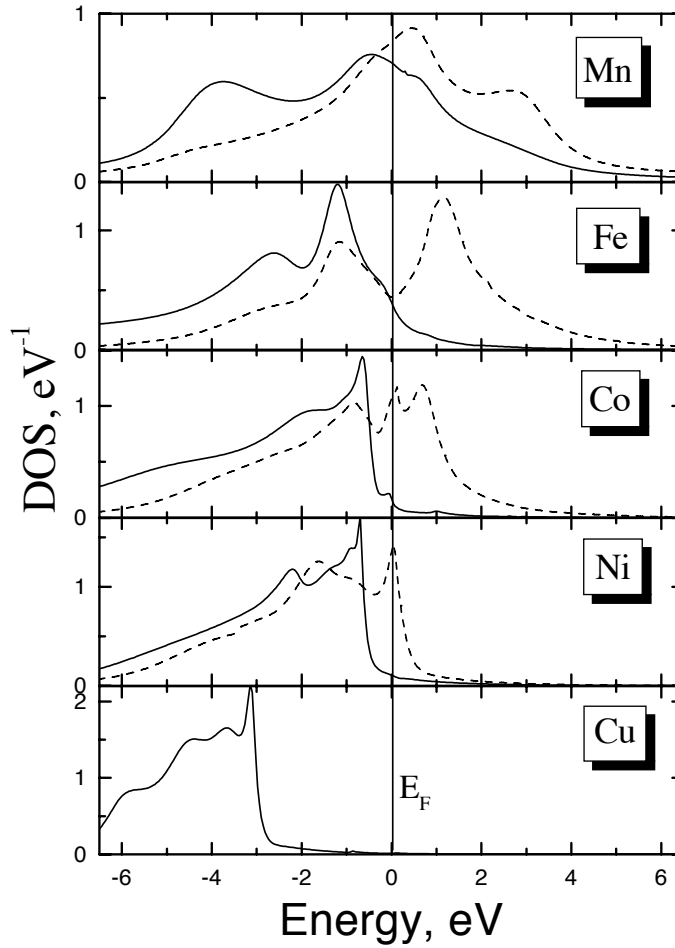


Fig. 12: Spin-polarized partial 3d density of states for different transition metals at temperature $T=750$ K. The full line is the spin-up, the dashed line the spin-down DOS.

the maxima of the spectral density (dark regions) coincide with the experimentally obtained band structure. However, for occupied majority spin states at about -3 eV the distribution of the spectral density is rather broad and the description of these states in terms of the quasi-particle dispersion is problematic. This conclusion is in complete quantitative agreement with raw experimental data on angle-resolved spin-polarized photoemission [65] with the broad non-dispersive second peak in the spin-up spectral function around -3 eV.

Comparison of the DOS for transition metals in the Fig. 12 shows interesting correlation effects. First of all, the most prominent difference from the LDA calculation is observed for antiferromagnetic *fcc*-Mn. There is the clear formation of lower and upper Hubbard bands around ± 3 eV. Such behavior is related with the half-filled Mn *d*-shell, which corresponds to a large phase space for particle-hole fluctuations. For ferromagnetic *bcc*-Fe, the *p-h* excitations are suppressed by the large exchange splitting and a *bcc* structural minimum in the DOS near the Fermi level. In the case of ferromagnetic *fcc*-Co and Ni, correlation effects are more important than for Fe since there is no structural *bcc*-dip in the density of states. One can see the formation of a "three-peak" structure for the spin-down DOS for Co and Ni and satellite formation around

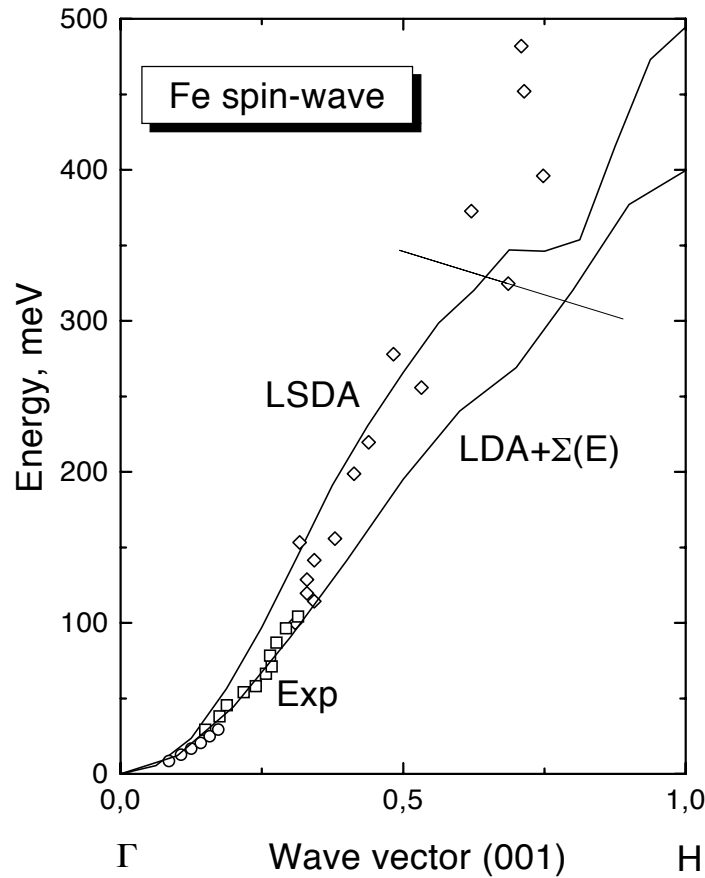


Fig. 13: The spin-wave spectrum for ferromagnetic iron in LDA and LDA+ Σ compared with different experiments (circles [16], squares [17], and diamonds [18]) (a); The corresponding spin-wave spectrum from the LDA+ Σ scheme in the (110) plane (b).

-5 eV. In order to describe the satellite formation more carefully one needs to include T -matrix effects [29, 34] or use the QMC scheme in LDA+DMFT calculations. Finally, there are no big correlation effects in non-magnetic *fcc*-Cu since the *d*-states are located well below the Fermi level.

Using the self-consistent values for $\Sigma_m(i\omega)$ computed by QMC, we calculate the exchange interactions (Eq. 56) and spin-wave spectrum (Eq. 57) using the four-dimensional fast Fourier transform (FFT) method [66] in $(\mathbf{k}, i\omega)$ space with a mesh of $20^3 \times 320$. The spin-wave spectrum for ferromagnetic iron is presented in Fig. 13 with comparison to the results of LDA-exchange calculations [57] and with different experimental data [67–69]. The room-temperature neutron scattering experiments have a sample dependence (Fe-12%Si in Ref. [67, 69] and Fe-4%Si in Ref. [68]) due to problems with the *bcc*-Fe crystal growth. Note that for high-energy spin-waves the experimental data [69] has large error-bars due to Stoner damping (we show one experimental point with the uncertainties in \mathbf{q} space). On the other hand, the expression of magnon frequency in terms of exchange parameters itself becomes problematic in that region due to the breakdown of the adiabatic approximation. Therefore we think that the comparison of our theoretical results with the experimental spin-wave spectrum for large energy needs addi-

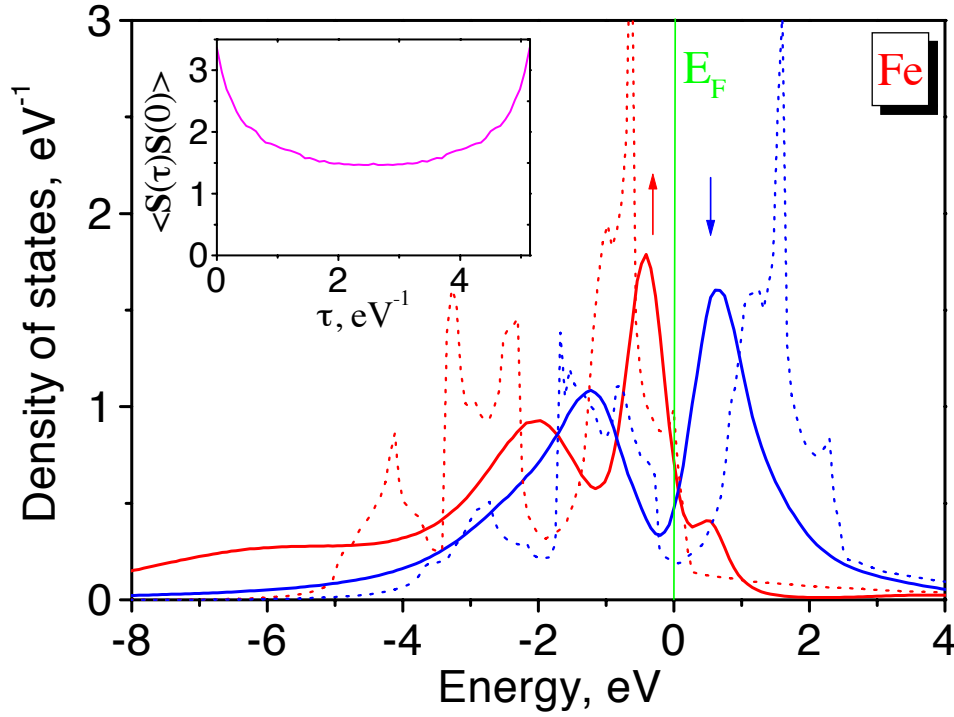


Fig. 14: LDA+DMFT results for ferromagnetic iron ($T = 0.8 T_C$). The partial densities of the d -states (full lines) are compared with the corresponding LDA results at zero temperature (dashed lines) for the spin-up (red lines, arrow-up) and spin-down (blue lines, arrow-down) states. The insert shows the spin-spin autocorrelation function for $T=1.2 T_C$.

tional investigation of Stoner excitations and calculations of the dynamical susceptibility in the LDA+DMFT approach [43]. Within the LDA scheme one could use the linear-response formalism [70] to calculate the spin-wave spectrum with the Stoner renormalizations, which should give in principle the same spin-wave stiffness as our LDA calculations. Our LDA spin-wave spectrum agrees well with the results of frozen magnon calculations [71, 72].

At the lower-energy, where the present adiabatic theory is reliable, the LDA+DMFT spin-wave spectrum agrees better with experiments than the result of the LDA calculations. Experimental values of the spin-wave stiffness $D = 280 \text{ meV/\AA}^2$ [68] agree well with the theoretical LDA+DMFT estimate of 260 meV/\AA^2 [38].

Self-consistent LDA+DMFT results for the local spectral function of iron and nickel are shown in Figs. 14 and 15, respectively. The LDA+DMFT approach describes well all the qualitative features of the density of states (DOS), which is especially non-trivial for nickel. Our QMC results reproduce well the three main correlation effects on the one particle spectra below T_C [75–77]: the presence of the famous 6 eV satellite, the 30% narrowing of the occupied part of d -band and the 50% decrease of exchange splitting compared to the LDA results. Note that the satellite in Ni has substantially more spin-up contributions, in agreement with photoemission spectra [77]. The exchange splitting of the d -band depends very weakly on temperature from $T = 0.6 T_C$ to $T = 0.9 T_C$. Correlation effects in Fe are less pronounced than in Ni, due to its large spin-splitting and the characteristic *bcc*-structural dip in the density of states for spin-down

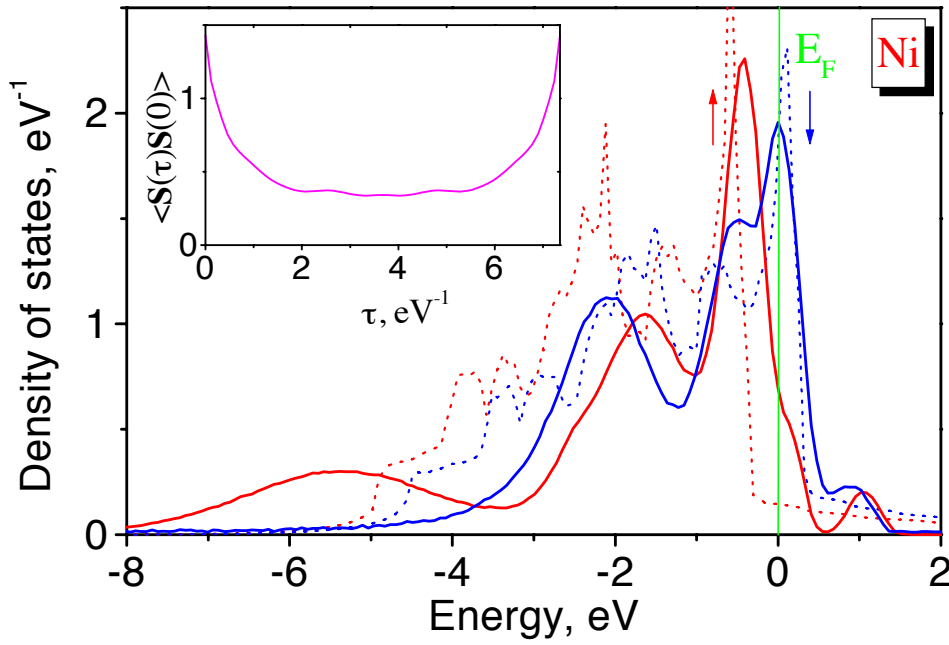


Fig. 15: LDA+DMFT results for ferromagnetic nickel ($T = 0.9 T_C$). The partial densities of the d -states (full lines) is compared with the corresponding LDA results at zero temperature (dashed lines) for the spin-up (red lines, arrow-up) and spin-down (blue lines, arrow-down) states. The insert shows the spin-spin autocorrelation function for $T = 1.8 T_C$.

states near the Fermi level that reduces the density of states for particle hole excitations.

Now we discuss the application of the LDA+DMFT approach to the description of finite-temperature magnetic properties of iron and nickel. While density functional theory can, in principle, provide a rigorous description of the thermodynamic properties, at present there is no accurate practical implementation available. As a result the finite-temperature properties of magnetic materials are estimated following a simple suggestion [57], whereby constrained DFT at $T = 0$ is used to extract exchange constants for a *classical* Heisenberg model, which in turn is solved using approximation methods (e.g., RPA, mean field) from classical statistical mechanics of spin systems [57, 78–80]. The most recent implementation of this approach gives good values for the transition temperature of iron but not of nickel [81]. While these localized spin models give, by construction, at high temperatures a Curie-Weiss like magnetic susceptibility, as observed experimentally in Fe and Ni, they encountered difficulties in predicting the correct values of the Curie constants [82].

The uniform spin susceptibility in the paramagnetic state: $\chi_{q=0} = dM/dH$ was extracted from QMC simulations by measuring the induced magnetic moment in a small external magnetic field. The dynamical mean-field results account for the Curie-Weiss law which is observed experimentally in Fe and Ni. As the temperature increases above T_C , the atomic character of the system is partially restored resulting in an atomic-like susceptibility with effective moment

$$\chi_{q=0} = \frac{\mu_{\text{eff}}^2}{3(T - T_C)} \quad (59)$$

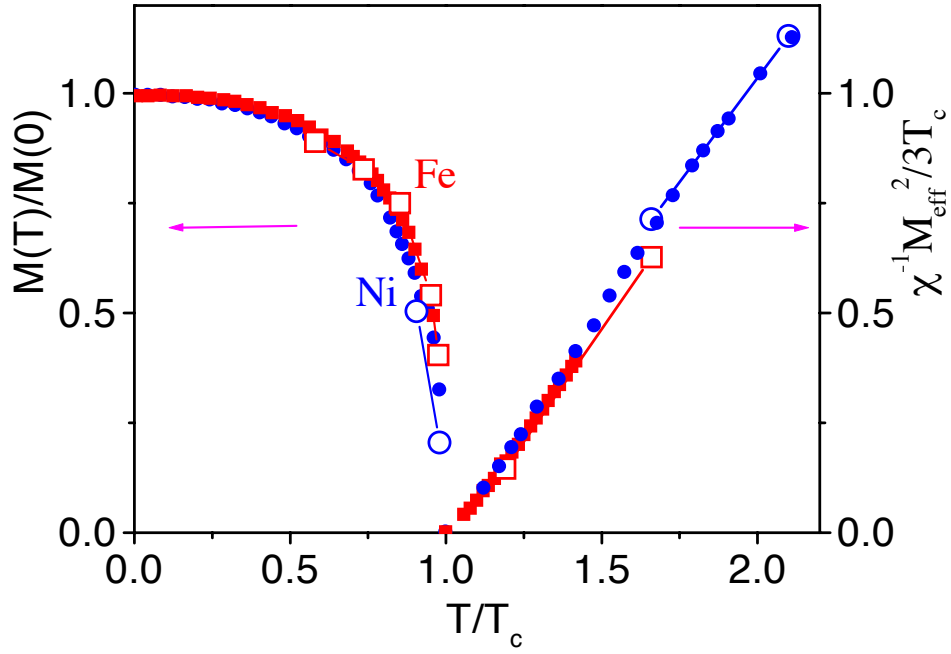


Fig. 16: Temperature dependence of ordered moment and inverse ferromagnetic susceptibility for Fe (open squares) and Ni (open circles) compared with experimental results (filled symbols).

The temperature dependence of the ordered magnetic moment below the Curie temperature and the inverse of the uniform susceptibility above the Curie point are plotted in Fig. 16 together with the corresponding experimental data for iron and nickel [83]. The LDA+DMFT calculations describe the magnetization-curve and the slope of the high-temperature Curie-Weiss susceptibility remarkably well. The calculated values of high-temperature magnetic moments extracted from the uniform spin susceptibility are $\mu_{\text{eff}} = 3.09$ (1.50) μ_B for Fe (Ni), in good agreement with the experimental data $\mu_{\text{eff}} = 3.13$ (1.62) μ_B for Fe (Ni) [83].

One can estimate the value of the Curie temperature of Fe and Ni from the disappearance of spin polarization in the self-consistent DMFT solution and from the Curie-Weiss law in Eq. (59). Our estimates $T_C = 1900$ (700) K are in reasonable agreement with experimental values of 1043 (631) K for Fe (Ni) respectively [83], considering the single-site nature of the DMFT approach, which is not able to capture the reduction of T_C due to long wavelength spin waves. These effects are governed by the spin wave stiffness. Since the ratio T_C/a^2D of the spin wave stiffness D to T_C is nearly a factor of 3 larger for Fe than for Ni [83] (a is the lattice spacing), we expect the Curie temperature from DMFT to be much higher than the observed T_C in Fe than in Ni. Note that this is a consequence of the long-range oscillating character of exchange interactions in iron compared to short-range ferromagnetic exchange interactions in nickel [81]. Quantitative calculations demonstrating the sizable reduction of T_C due to spin waves in Fe in the framework of a Heisenberg model were performed in Ref [81]. Moreover including additional spin-flip terms in the rotationally invariant Coulomb matrix will drastically reduce the effective Curie temperature for the case of iron with its approximately two unpaired electrons [88].

Within dynamical mean-field theory one can also compute the local spin susceptibility

$$\chi_{\text{loc}} = \frac{g_s^2}{3} \int_0^\beta d\tau \langle \mathbf{S}(\tau) \mathbf{S}(0) \rangle, \quad (60)$$

where $g_s = 2$ is the gyromagnetic ratio, $\mathbf{S} = \frac{1}{2} \sum_{m,\sigma,\sigma'} c_{m\sigma}^\dagger \vec{\sigma}_{\sigma\sigma'} c_{m\sigma'}$ the single-site spin operator, and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices. It differs from the $\mathbf{q} = \mathbf{0}$ susceptibility by the absence of spin polarization in the Weiss field of the impurity model. Eq. (60) cannot be probed directly in experiments but it is easily computed in DMFT-QMC. Its behavior as a function of temperature gives a very intuitive picture of the degree of correlations in the system. In a weakly correlated system we expect Eq. (60) to be nearly temperature-independent, while in a strongly correlated system we expect a leading Curie-Weiss behavior at high temperatures $\chi_{\text{loc}} = \mu_{\text{loc}}^2 / (3T + C)$, where μ_{loc} is an effective local magnetic moment. In the Heisenberg model with spin S , $\mu_{\text{loc}}^2 = S(S+1)g_s^2$ and for well-defined local magnetic moments (e.g., for rare-earth magnets) this quantity should be temperature-independent. For the itinerant electron magnets μ_{loc} is temperature-dependent, due to a variety of competing many body effects such as Kondo screening and the reduction of local magnetic moment by temperature [27]. All these effects are included in the DMFT calculations. The τ -dependence of the correlation function $\langle \mathbf{S}(\tau) \mathbf{S}(0) \rangle$ results in the temperature dependence of μ_{loc} and is displayed in the inserts on the Figs. 14,15. Iron can be considered as a magnet with very well-defined local moments above T_C (the τ -dependence of the correlation function is relatively weak), whereas nickel is a more itinerant electron magnet (stronger τ -dependence of the local spin-spin autocorrelation function).

The comparison of the values of the local and the $\mathbf{q} = \mathbf{0}$ susceptibility gives a crude measure of the degree of short range order which is present above T_C . As expected, the moments extracted from the local susceptibility, Eq. (60), are a bit smaller ($2.8 \mu_B$ for iron and $1.3 \mu_B$ for nickel) than those extracted from the uniform magnetic susceptibility. This reflects the small degree of short-range correlations that remains well above T_C [85]. The high-temperature LDA+DMFT data clearly show the presence of a local-moment above T_C . This moment is correlated with the presence of high energy features (of the order of the Coulomb energies) in the photoemission. This is also true below T_C , where the spin dependence of the spectra is more pronounced for the satellite region in nickel than for the quasiparticle bands near the Fermi level (Fig. 15). This can explain the apparent discrepancies between different experimental determinations of the high-temperature magnetic splittings [84, 86, 87] as being the result of probing different energy regions. The resonant photoemission experiments [86] reflect the presence of local-moment polarization in the high-energy spectrum above T_C in nickel, while the low-energy ARPES investigations [87] result in non-magnetic bands near the Fermi level. This is exactly the DMFT view on the electronic structure of transition metals above T_C . Fluctuating moments and atomic-like configurations are large at short times, which results in correlation effects in the high-energy spectra such as spin-multiplet splittings. The moment is reduced at longer time scales, corresponding to a more band-like, less correlated electronic structure near the Fermi level.

8 Conclusions

We have discussed a general scheme for investigating the magnetic properties for correlated itinerant-electron systems. This approach is based on the combination of the first-principle LDA scheme with dynamical mean-field theory. The application of the LDA+DMFT method gives an adequate description of the quasiparticle electronic structure of ferromagnetic transition metals. The main correlation effects in the electron energy spectrum are a strong damping of the occupied states more than 1 eV from the Fermi level E_F and essentially a depolarization of the states in the vicinity of E_F . We obtain a reasonable agreement with different experimental spectral data (spin-polarized photo- and thermo-emission). The method is quite universal and can be applied for other correlated d - and f -electron magnetic systems.

We discussed as well a general method for the investigation of magnetic interactions in correlated electron systems. Our general expressions are also valid in the relativistic case and can be used for the calculation of both exchange and Dzyaloshinskii-Moriya interactions, and magnetic anisotropy [38]. The illustrative example of ferromagnetic iron shows that correlation effects in the exchange interactions may be noticeable even in such moderately correlated systems. For rare-earth metals and their compounds, colossal magnetoresistance materials, or high- T_c systems, this effect may be even more important.

This work demonstrates an essential difference between the spin density functional and the LDA+DMFT approach. The latter deals with the thermodynamic potential as a functional of the local Green function rather than the electron density. Nevertheless, there is a close connection between the two techniques (the self-energy corresponds to the exchange-correlation potential). In particular, an analog of the local force theorem can be proved within the LDA+DMFT approach. It may be useful not only for the calculation of magnetic interactions but also for elastic stresses, in particular pressure, and other physical properties of correlated magnetic systems.

References

- [1] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)
- [2] A.R. Mackintosh, O.K. Andersen, in M. Springford (ed.): *Electrons at the Fermi Surface* (Cambridge University Press, 1980) p. 145
- [3] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, *Rev. Mod. Phys.* **78**, 865 (2006)
- [4] A.B. Migdal, *Theory of finite Fermi Systems and applications to atomic nuclei* (Interscience Publishers, New York, 1967)
- [5] P. Nozières, *Theory of interacting Fermi systems* (Benjamin, New York, 1964)
- [6] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964);
W. Kohn and L.J. Sham, *ibid.* **140**, A1133 (1965);
R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989)
- [7] N.F. Mott, *Metal-Insulator Transitions* (Taylor and Francis, London 1974)
- [8] D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980)
- [9] K. Terakura, A.R. Williams, T. Oguchi, and J. Kubler, *Phys. Rev. Lett.* **52**, 1830 (1984)
- [10] W.E. Pickett, *Rev. Mod. Phys.* **61**, 433 (1989)
- [11] V.I. Anisimov, F. Aryasetiawan, and A.I. Lichtenstein, *J. Phys.: Condens. Matter* **9**, 767 (1997)
- [12] P. Fulde, J. Keller, and G. Zwicknagl, *Solid State Phys.* **41**, 1 (1988)
- [13] D.E. Eastman, F.J. Himpsel, and J.A. Knapp, *Phys. Rev. Lett.* **44**, 95 (1980)
- [14] J. Staunton, B.L. Gyorffy, A.J. Pindor, G.M. Stocks, and H. Winter, *J. Phys. F* **15**, 1387 (1985)
- [15] P. Genoud, A.A. Manuel, E. Walker, and M. Peter, *J. Phys.: Cond. Matter* **3**, 4201 (1991)
- [16] A. Vaterlaus, F. Milani, and F. Meier *Phys. Rev. Lett.* **65**, 3041 (1990)
- [17] R. Monnier, M.M. Steiner, and L.J. Sham, *Phys. Rev. B* **44**, 13678 (1991)
- [18] M.I. Katsnelson and A.I. Lichtenstein, *J. Phys.: Cond. Matter* **11**, 1037 (1999)
- [19] D. Chanderis, J. Lecante, and Y. Petroff, *Phys. Rev. B* **27**, 2630 (1983);
A. Gutierrez and M.F. Lopez, *Phys. Rev. B* **56**, 1111 (1997)

- [20] R.E. Kirby, B. Kisker, F.K. King, and E.L. Garwin, Solid State Commun. **56**, 425 (1985)
- [21] L. Hedin, Phys. Rev. **139**, A796 (1965);
F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998)
- [22] E.K.U. Gross and W. Kohn, Adv. Quant. Chem. **21**, 255 (1990)
- [23] A.A. Abrikosov, L.P. Gorkov, and I.E. Dzialoshinskii:
Methods of Quantum Field Theory in Statistical Physics (Pergamon, New York 1965);
G. Mahan, *Many-Particle Physics* (Plenum, New York 1981)
- [24] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960);
see also G.M. Carneiro and C.J. Pethick, Phys. Rev. B **11**, 1106 (1975)
- [25] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 289 (1961);
G. Baym, Phys. Rev. **127**, 1391 (1962)
- [26] T. Jarlborg, Rep. Prog. Phys. **60**, 1305 (1997)
- [27] T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Berlin 1985)
- [28] T. Kotani, Phys. Rev. B **50**, 14816 (1994); T. Kotani and H. Akai, Phys. Rev. B **54**, 16502 (1996); N.E. Zein, V.P. Antropov, and B.N. Harmon, J. Appl. Phys. **87**, 5079 (2000)
- [29] A. Liebsch, Phys. Rev. Lett. **43**, 1431 (1979); Phys. Rev. B **23**, 5203 (1981)
- [30] M.M. Steiner, R.C. Albers, and L.J. Sham, Phys. Rev. B **45**, 13272 (1992)
- [31] V. Yu. Irkhin, M.I. Katsnelson, and A.V. Trefilov, J. Phys. : Cond. Matter **5**, 8763 (1993);
S.V. Vonsovsky, M.I. Katsnelson, and A.V. Trefilov,
Phys. Metals Metallography **76**, 247, 343 (1993)
- [32] T. Greber, T.J. Kreuntz, and J. Osterwalder, Phys. Rev. Lett. **79**, 4465 (1997);
B. Sinkovich, L.H. Tjeng, N.B. Brooks, J. B. Goedkoop, R. Hesper, E. Pellegrin,
F.M.F. de Groot, S. Altieri, S.L. Hulbert, E. Shekel, and G. A. Sawatzky,
Phys. Rev. Lett. **79** 3510 (1997)
- [33] G. Treglia, F. Ducastelle, and D. Spanjaard, J. Phys. (Paris) **43** 341 (1982)
- [34] V. Drchal, V. Janiš, and J. Kudrnovský in A. Gonis, N. Kioussis, and M. Ciftan (eds.):
Electron Correlations and Material Properties(Kluwer/Plenum, New York 1999) p. 273.
- [35] J. Igarashi, J. Phys. Soc. Japan **52**, 2827 (1983);
F. Manghi, V. Bellini, and C. Arcangeli, Phys. Rev. B **56**, 7149 (1997)
- [36] W. Nolting, S. Rex, and S. Mathi Jaya, J. Phys.: Cond. Matter **9**, 1301 (1987)
- [37] A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **57**, 6884 (1998)

- [38] M.I. Katsnelson and A.I. Lichtenstein, Phys. Rev. B **61**, 8906 (2000)
- [39] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, J. Phys. Cond. Matter **9**, 7359 (1997)
- [40] A.I. Liechtenstein, V.I. Anisimov, and J. Zaanen, Phys. Rev. B **52**, R5467 (1995)
- [41] J. Hubbard, Proc. Roy. Soc., **A276**, 238 (1963); **A277**, 237 (1964)
- [42] O.K. Andersen, Phys. Rev. B **12**, 3060 (1975);
O.K. Andersen and O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984)
- [43] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [44] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
- [45] M.J. Rozenberg, Phys. Rev. B **55**, R4855 (1997);
K. Takegahara, J. Phys. Soc. Japan **62**, 1736 (1992)
- [46] J.E. Hirsch and R. M. Fye, Phys. Rev. Lett. **25**, 2521 (1986)
- [47] M. Jarrell and J.E. Gubernatis, Physics Reports **269**, 133 (1996)
- [48] A.N. Rubtsov and A.I. Lichtenstein, JETP Lett. **80**, 61 (2004)
- [49] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)
- [50] P. Werner, A. Comanac, L. de Medici, M. Troyer, and A.J. Millis, Phys. Rev. Lett. **97**, 076405 (2006)
- [51] P. Werner and A.J. Millis, Phys. Rev. B **74**, 155107 (2006)
- [52] K. Haule, Phys. Rev. B **75**, 155113 (2007)
- [53] N.E. Bickers and D.J. Scalapino, Ann. Phys. (N. Y.) **193**, 206 (1989)
- [54] G. Esirgren and N.E. Bickers, Phys. Rev. B **57**, 5376 (1998)
- [55] M. Fleck, A.I. Liechtenstein, A.M. Oles, L. Hedin, and V.I. Anisimov, Phys. Rev. Lett. **80**, 2393 (1998)
- [56] A.R. Mackintosh, O.K. Andersen, in M. Springford (ed.): *Electrons at the Fermi Surface* (Cambridge University Press, 1980) p. 145
- [57] A.I. Liechtenstein, M.I. Katsnelson, and V.A. Gubanov, J. Phys. F **14**, L125;
Solid State Commun. **54**, 327 (1985); A.I. Liechtenstein, M.I. Katsnelson, V.P. Antropov,
and V.A. Gubanov, J. Magn. Magn. Mater. **67**, 65 (1987)

- [58] V.P. Antropov, M.I. Katsnelson, and A.I. Lichtenstein, *Physica B* **237-238**, 336 (1997)
- [59] M.I. Katsnelson, and A.I. Lichtenstein, *Eur. Phys. J. B* **30**, 9 (2002)
- [60] V.I. Anisimov and O. Gunnarsson, *Phys. Rev. B* **43**, 7570 (1991)
- [61] A.M. Oles and G. Stolhoff, *Phys. Rev. B* **29**, 314 (1984)
- [62] V.I. Anisimov, I.V. Solovjev, M.A. Korotin, M.T. Czyzyk, and G.A. Sawatzky, *Phys. Rev. B* **48**, 16929 (1993)
- [63] H.J. Vidberg and J.W. Serene, *J. Low Temp. Phys.* **29**, 179 (1977)
- [64] V. Yu. Irkhin and M.I. Katsnelson, *Physics-Uspekhi* **37**, 659 (1994)
- [65] E. Kisker, K. Schroeder, T. Gudat, and M. Campagna, *Phys. Rev. B* **31**, 329 (1985)
- [66] S. Goedecker, *Comp. Phys. Commun.* **76**, 294 (1993)
- [67] J.W. Lynn, *Phys. Rev. B* **11**, 2624 (1975)
- [68] H.A. Mook and R.M. Nicklow, *Phys. Rev. B* **7**, 336 (1973)
- [69] T.G. Peerring, A.T. Boothroyd, D.M. Paul, A.D. Taylor, R. Osborn, R.J. Newport, H.A. Mook, *J. Appl. Phys.* **69**, 6219 (1991)
- [70] S.Y. Savrasov, *Phys. Rev. Lett.* **81**, 2570 (1998)
- [71] L.M. Sandratskii and J. Kübler, *J. Phys.: Condens. Matter.* **4**, 6927 (1992)
- [72] S.V. Halilov, H. Eschrig, A.Y. Perlov, and P.M. Oppeneer, *Phys. Rev. B* **58**, 293 (1998)
- [73] J.A. Hertz and D.M. Edwards, *J. Phys. F* **3**, 2174 (1973)
- [74] I.V. Solovjev and K. Terakura, *Phys. Rev. B* **58**, 15496 (1998)
- [75] M. Iwan, F.J. Himpsel, and D.E. Eastman, *Phys. Rev. Lett.* **43**, 1829 (1979)
- [76] W. Eberhardt and E.W. Plummer, *Phys. Rev. B* **21**, 3245 (1980)
- [77] K.N. Altmann, D.Y. Petrovykh, G.J. Mankey, N. Shannon, N. Gilman, M. Hochstrasser, R.F. Willis, and F.J. Himpsel, *Phys. Rev. B* **61**, 15661 (2000)
- [78] N.M. Rosengaard and B. Johansson, *Phys. Rev. B* **55**, 14975 (1997)
- [79] S.V. Halilov, H. Eschrig, A.Y. Perlov, and P. M. Oppeneer, *Phys. Rev. B* **58**, 293 (1998)
- [80] V.P. Antropov, M.I. Katsnelson, B.N. Harmon, M. van Schilfgaarde, and D. Kusnezov, *Phys. Rev. B* **54**, 1019 (1996)

- [81] M. Pajda, J. Kudrnovsky, I. Turek, V. Drchal, and P. Bruno, Phys. Rev. B **64**, 174402 (2001)
- [82] J.B. Staunton and B.L. Gyorffy, Phys. Rev. Lett. **69**, 371 (1992)
- [83] E.P. Wolfarth (de.): *Ferromagnetic materials*, vol. 1 (North-Holland, Amsterdam, 1986)
- [84] E. Kisker, K. Schröder, M. Campagna, and W. Gudat , Phys. Rev. Lett. **52**, 2285 (1984);
A. Kakizaki, J. Fujii, K. Shimada, A. Kamata, K. Ono, K.H. Park, T. Kinoshita, T. Ishii,
and H. Fukutani, Phys. Rev. Lett. **72**, 2781 (1994)
- [85] H.A. Mook and J.W. Lynn, J. Appl. Phys. **57**, 3006 (1985)
- [86] B. Sinkovic, L.H. Tjeng, N.B. Brookes, J.B. Goedkoop, R. Hesper, E. Pellegrin, F.M.F. de
Groot, S. Altieri, S.L. Hulbert, E. Shekel, and G. A. Sawatzky,
Phys. Rev. Lett. **79**, 3510 (1997)
- [87] T.J. Kreutz, T. Greber, P. Aebi, and J. Osterwalder, Phys. Rev. B **58**, 1300 (1998)
- [88] A.S. Belozarov, I. Leonov, V.I. Anisimov, Phys. Rev. B **87**, 125138 (2013)

6 Monte Carlo Methods with Applications to Spin Systems

Werner Krauth

Laboratoire de physique statistique

Ecole normale supérieure

24 rue Lhomond, 75005 Paris, France

Contents

1	Introduction	2
2	Markov chains and incompressible flow	2
3	A priori probabilities	3
4	Local algorithm, faster-than-the-clock algorithms	4
5	Cluster algorithms	6
6	Coupling approach	7

1 Introduction

In my lecture at the Jülich Autumn School 2013, I discuss Monte Carlo methods, and their application to spin systems, in the language of my recent text book [1], but in much less detail. During the lecture, I first concentrate on the relation between Monte Carlo methods and the concept of incompressible flow (Section 2), which allows one to understand the crucial detailed and global balance conditions. I then introduce to the “a priori” probabilities, which are at the basis of the Metropolis-Hastings algorithm (Section 3). After an introduction to the local algorithms (heatbath, local Metropolis), and the classic faster-than-the-clock approaches (Section 4), I then discuss the Swendsen-Wang and Wolff cluster algorithms for spin systems (Section 5), as an illustration of the great liberty one has in designing Monte Carlo algorithms. I finish with the coupling approach (Section 6) which has provided essential for rigorously understanding for how long a simulation must be run until the computational output reflects thermal equilibrium (the classical Boltzmann distribution or the quantum density matrix) rather than the initial state of the Markov chain.

2 Markov chains and incompressible flow

Markov-chain Monte Carlo strives to sample a distribution π by starting from another distribution π_0 and using an incremental algorithm $p_{i \rightarrow j}$ to move between configurations i and j . In each time step (iteration) $t = 1, 2, 3, \dots$, the probability distribution π^t is sampled. The target distribution is approached in the limit of infinite simulation time $\pi = \pi^\infty$. Convergence is always exponential in time, and it can be assured as long as we satisfy (in addition to an ergodicity requirement) the global balance condition. This means that the flow into each configuration i equals the flow out of it:

$$\sum_k P_{i \rightarrow k} = \sum_l P_{l \rightarrow i} \quad (\text{global balance}). \quad (1)$$

Here, the flow $P_{k \rightarrow i}$ denotes the product of the stationary probability π_k with the conditional probability $p_{k \rightarrow i}$ to move from k to i given that the system is at k . The global flow condition can be easily understood under the stationarity condition, as the first sum in Eq. (1) equals $\sum_k P_{i \rightarrow k} = \pi_i$ (see [1], Section 1.1.4). Eq. (1) is satisfied in particular if we balance the flow between each pair of configurations i and k individually:

$$P_{i \rightarrow k} = P_{k \rightarrow i} \quad \forall i, k \quad (\text{detailed balance}). \quad (2)$$

This detailed balance condition is satisfied for example by the Metropolis algorithm

$$P_{i \rightarrow k} = \min(\pi_i, \pi_k) \quad (3)$$

(the right-hand-side of this equation is manifestly symmetric in i and k , so is the left-hand-side). Usually, Eq. (3) is written in terms of the probabilities $p_{i \rightarrow k}$, etc, that is by dividing by $\pi_i > 0$. This yields

$$p_{i \rightarrow k} = \min(1, \pi_k / \pi_i), \quad (4)$$

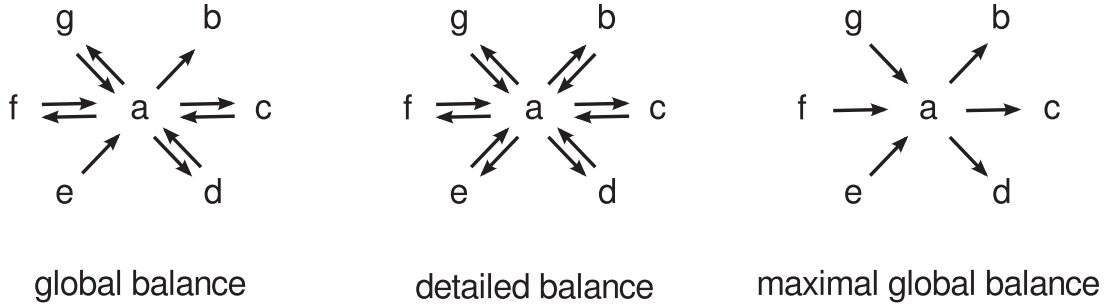


Fig. 1: Balance condition for Markov-chain Monte Carlo moves. The global balance condition (left) is necessary for the convergence of the Markov chain towards the stationary probability measure π . The time-honored detailed balance condition (center) is often realized through the Metropolis algorithm of Eq. (3) and Eq. (4). The unidirectional motion of the maximal global balance condition (right) is realized in a number of modern algorithms (see [2]. Figure adapted from [3]).

an equation that has been written down in thousands of papers and a huge number of computer programs. There are many choices of P which satisfy these balance conditions. This implies that a great many valid Monte Carlo algorithms are out there, some of them slow and some of them fast, most of the latter ones yet to be discovered.

Much more could be said, and has been written, about the detailed and global balance conditions, but the most important finding is the following: While convergence towards the equilibrium distribution π is generally exponential ($\propto \exp(-t/\tau)$), even for continuum systems, it is very difficult to estimate the correlation time τ . This question — what is the correlation time — is one of the three great conceptual issues in Monte Carlo algorithms and, really, the most serious one. The other two are finite-size scaling, and the estimation of the variance for observables (see [1], Section 1.4). We will take up the convergence issue in Section 6, in the context of coupling algorithms.

3 A priori probabilities

How do we move from configuration i to configuration j , and what do we mean with the (conditional) probability $p_{i \rightarrow j}$ to go from configuration i to configuration j (see our discussion in Section 1)? This question is best discussed with the help of *a priori probabilities*: While the flow $P_{i \rightarrow j}$ is a product of the stationary probability π_i and the conditional probability $p_{i \rightarrow k}$, the latter is put together from a probability to *propose* the move from i to j (the “a priori” probability) times the conditional probability to *accept* it:

$$p_{i \rightarrow j} = \mathcal{A}_{i \rightarrow j} p_{i \rightarrow j}^{\text{acc}}. \quad (5)$$

This equation can be illustrated by a simple example, as the “triangle algorithm” for a particle in a box (see [1], Sect. 1.1.6). As the balance condition of Eq. (1) only considers the flow, we are free to choose an arbitrary a priori probability \mathcal{A} and bias it with a given acceptance

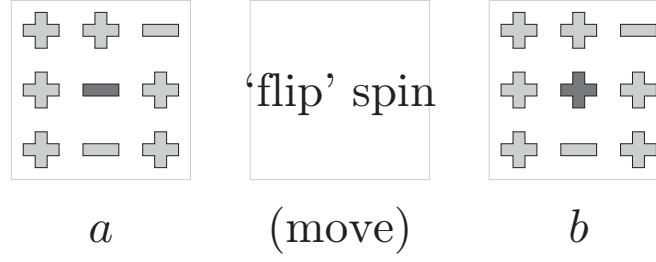


Fig. 2: Local Monte Carlo move $a \rightarrow b$ in the Ising model. The spin to be flipped is sampled uniformly through the lattice and the flip should be accepted with probability $\min(1, \exp(-\beta(E_b - E_a)))$ (from [1]).

probability in order to fall back on our feet (obtain the correct flow). With a priori probabilities, the Metropolis Monte Carlo algorithm takes the form

$$p_{i \rightarrow j}^{\text{acc}} = \min \left(1, \frac{\mathcal{A}_{j \rightarrow i}}{\pi_i} \frac{\pi_j}{\mathcal{A}_{i \rightarrow j}} \right). \quad (6)$$

If the a priori probability $\mathcal{A}_{i \rightarrow j}$ to propose the move from i to j is equal to π_j , the acceptance probability $p_{i \rightarrow j}^{\text{acc}}$ is equal to one and, in a sense, we have already solved the problem. This means two things: (i) we have no rejections and (ii) we can usually make quite large moves, as the reason for making small ones was the presence of rejections. See [1], Section 1.1.6, for a discussion of the relation between this Metropolis-Hastings algorithm and perturbation theory of quantum physics.

4 Local algorithm, faster-than-the-clock algorithms

Let us now do our first real Monte Carlo simulation of Ising spins $\sigma_k = \pm 1, k = 1 \dots N$ on a lattice, say the two-dimensional square lattice. We use the energy

$$E = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j, \quad (7)$$

where i and j are neighbors on the lattice. In the following, we choose $J = 1$, and the stationary probability of a configuration σ is given by $\pi_\sigma = \exp[-E(\sigma)/kT]$, where k is the Boltzmann constant and T the temperature, with $\beta = 1/kT$.

The Ising model can be simulated by the Metropolis algorithm, by uniformly sampling at each time step a site i , as in Fig. 2. If we call “ a ” the original configuration and “ b ” the configuration obtained through the flip $\sigma_i \rightarrow -\sigma_i$, then the Metropolis probability for accepting the move is given by

$$p = \min[1, \exp(-\beta(E_b - E_a))]. \quad (8)$$

One may picture the spin σ_i in the “molecular” field h_i equal to the sum of the neighboring spins, and the flipping probability turns out to be

$$p = \min[1, \exp(-2\beta\sigma_i h_i)]. \quad (9)$$

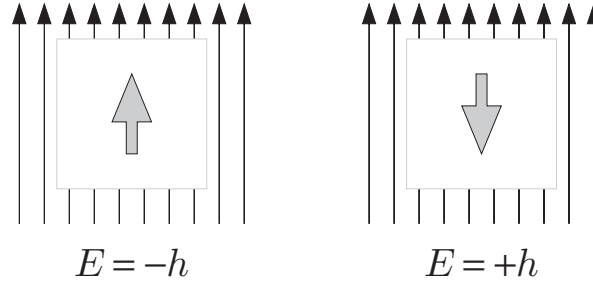


Fig. 3: A single spin in an external field: building block of local Monte Carlo algorithms (figure adapted from [1]).

Instead of the Metropolis algorithm, we may also use the heatbath algorithm, which consists in taking out the spin i and in replacing it with thermalized values of σ_i

$$\pi_h^+ = \frac{e^{-\beta E^+}}{e^{-\beta E^+} + e^{-\beta E^-}} = \frac{1}{1 + e^{-2\beta h}}, \quad (10)$$

$$\pi_h^- = \frac{e^{-\beta E^-}}{e^{-\beta E^+} + e^{-\beta E^-}} = \frac{1}{1 + e^{+2\beta h}}. \quad (11)$$

(See [1], Section 5.2.1 for simple programs, and the author's website for a basic Python program.) Practically, we sample a random number \mathcal{U} uniformly distributed in the unit interval. If $\mathcal{U} < \pi_h^+$, we set $\sigma_i = 1$, otherwise $\sigma_i = -1$). In the long run, the program will converge towards the Boltzmann distribution, that is, spin configurations σ will be visited with their Boltzmann probability π_σ .

In the local Metropolis or heatbath algorithm, the flip concerns a single spin subject to the molecular magnetic field produced by its neighbors. Let us analyze this problem, a single spin in a magnetic field, in more detail (see Fig. 3). At low temperature, the acceptance probability of the algorithm is very small: We consider the spin, and try to flip it (but reject the flip), then try again, reject again, etc. Many calculations will be wasted, as we compute probabilities and draw random numbers, but most often reject the move. There is a better way to organize the calculation: it consists in finding out how many times we will reject the flip before accepting it. This is called the faster-than-the-clock approach (in a single iteration, we may determine that the flip in Fig. 3 is rejected 7 times before it is accepted). As discussed in detail in [1], Section 7.2.2, the faster-than-the-clock approach is at the base of the classic Bortz-Kalos-Lebowitz (BKL) algorithm [4] among others. The general lesson is that very small acceptance probabilities, as they appear at low temperatures in the Ising model, can generally be avoided through a clever trick. The inverse is also true: in some recent approaches, almost all Monte Carlo move are accepted, and their results can be easily added up [5, 6]. We can then use the faster-than-the-clock approach to sample the first time when a long streak of events (or non-events) comes to an end.

5 Cluster algorithms

Close to the critical point, the transition between the paramagnetic and the ferromagnetic phases, local Monte Carlo algorithms slow down considerably. This is due to a complicated interplay between several factors: On the one hand, close to the critical point, finite-size effects are large and the physical observables at one system size N can be quite different from the ones for larger or smaller systems (in consequence, we must simulate very large systems, which is naturally CPU-time intensive). On the other hand, computing these observables for fixed N may take quite some CPU time because of critical slowing down. In a computational context, this phenomenon describes the appearance of slow variables (in the Ising model it is the total magnetization), whose distribution becomes very wide (at the critical point, total magnetizations from $-N$ through $+N$ appear with sizable probability). As the local Monte Carlo algorithm only flips a single spin at a time, it takes a long time to move across the probability distribution of the magnetization, and we experience slowing down (see [1], Section 5.2.1, for a detailed discussion).

To speed up the simulation, we should flip more than one spin concurrently, but this cannot be achieved by parallel flipping of uncorrelated spins: The attempted moves are indeed larger, but with the increased rejection rate, the effective move will get smaller instead of larger. We understand that we should propose larger moves but, quite importantly, we must get them accepted. The Wolff cluster algorithm [7] (an efficient version of the Swendsen-Wang cluster algorithm [8]) does this marvelously. It provides a prime example for the use of the Metropolis-Hastings paradigm. Our discussion closely mirrors [1], Section 5.2.3.

In the Wolff cluster algorithm, we start with one site, then add any of the neighboring sites with probability p , if they carry the same spin, etc. The construction in Fig. 4a then came to a halt because all of the 14 links $++$ across the cluster boundary were rejected. The move consists in flipping the cluster.

To apply Eq. (6), we compute the a priori probability $\mathcal{A}_{a \rightarrow b}$, the probability of stopping the cluster construction process at a given stage rather than continuing and including more sites (see the configuration a in Fig. 4). $\mathcal{A}_{a \rightarrow b}$ is given by an interior part (the two neighbors inside the cluster) and the stopping probability at the boundary: each site on the boundary of the cluster was once a pocket site and the construction came to a halt because none of the possible new edges was included. Precisely, the boundary $\partial\mathcal{C}$ of the cluster (with one spin inside and its neighbor outside) involves two types of edge:

$$\left\{ \begin{array}{l} \text{cluster in } a \\ \text{in Fig. 4} \end{array} \right\} : \overbrace{\left[\begin{array}{ccc} \text{inside} & \text{outside} & \# \\ + & - & n_1 \\ + & + & n_2 \end{array} \right]}^{\text{edges across } \partial\mathcal{C}} E|_{\partial\mathcal{C}} = n_1 - n_2 \quad (12)$$

(in Fig. 4, $n_1 = 17$, $n_2 = 14$). In order to acquire all information needed for Eq. (6), we must analyze likewise the return flip in Fig. 4 from configuration b back to configuration a . the same interior part as before, but a new boundary part $\mathcal{A}_{b \rightarrow a} = \mathcal{A}_{\text{in}} \cdot (1 - p)^{n_1}$, because

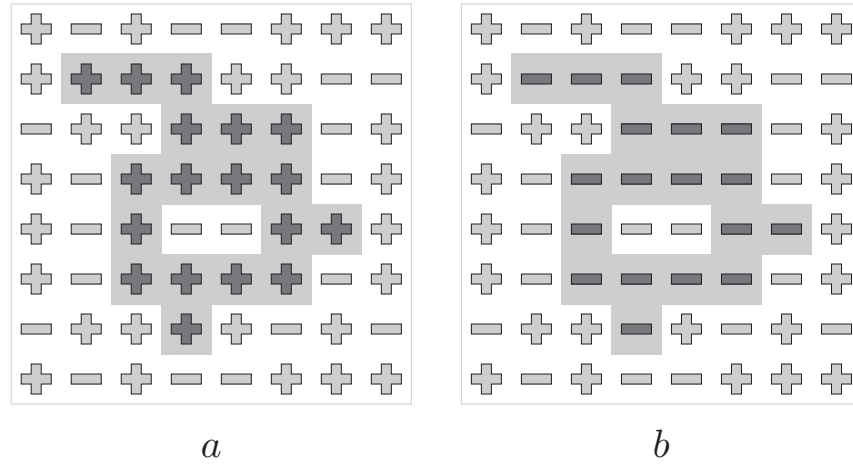


Fig. 4: Wolff cluster algorithm in the Ising model. Starting from an initial site, we move on to neighboring sites with the same value of the spin, and add them to the cluster with probability p . The cluster construction terminates, and the cluster is flipped, when no more neighboring sites can be added. In order to set up the algorithm correctly, we have to consider both the initial configuration a and the final configuration b (see Eq. (6)). For the special value of p given in Eq. (14), this algorithm is rejection free (figure taken from [1]).

there were n_1 opportunities to let the cluster grow and again none was accepted. We also note that $\pi_a = \pi_{\text{in}}\pi_{\text{out}} \exp(-\beta(n_1 - n_2))$, where π_{in} takes into account all the edges with both sites inside the cluster and π_{out} all those at its outside. By an argument similar to that above, the statistical weight of configuration b is $\pi_b = \pi_{\text{in}}\pi_{\text{out}} \exp(-\beta(n_2 - n_1))$. The interior and exterior contributions to the Boltzmann weight are the same as for configuration a . All the ingredients of the detailed-balance condition in Eq. (6) are now known:

$$e^{-\beta(n_1 - n_2)}(1 - p)^{n_2} p_{a \rightarrow b}^{\text{acc}} = e^{-\beta(n_2 - n_1)}(1 - p)^{n_1} p_{b \rightarrow a}^{\text{acc}}. \quad (13)$$

which leads to

$$p_{a \rightarrow b}^{\text{acc}} = \min \left[1, \left(\frac{e^{-2\beta}}{1 - p} \right)^{n_2} \left(\frac{1 - p}{e^{-2\beta}} \right)^{n_1} \right]. \quad (14)$$

The algorithm is at its peak efficiency for $p = 1 - \exp(-2\beta)$, where the acceptance probability is equal to one: we can construct the cluster, then flip it, build another one, turn it over This Wolff cluster algorithm [7], one of the first of its kind, can be implemented in a dozen or so lines of code: see the website of [1].

6 Coupling approach

As final subject, we now consider the problem of correlation times: that is we try to understand at what time the sampled configurations reflect the stationary distribution π^∞ and no longer the starting configuration π^0 . For many real-life problems, this complex and crucial question has no truly satisfactory solution: Although we know that the corrections to the equilibrium state decay as $e^{-t/\tau}$, we cannot really compute τ . Quite often whole communities underestimate τ ,

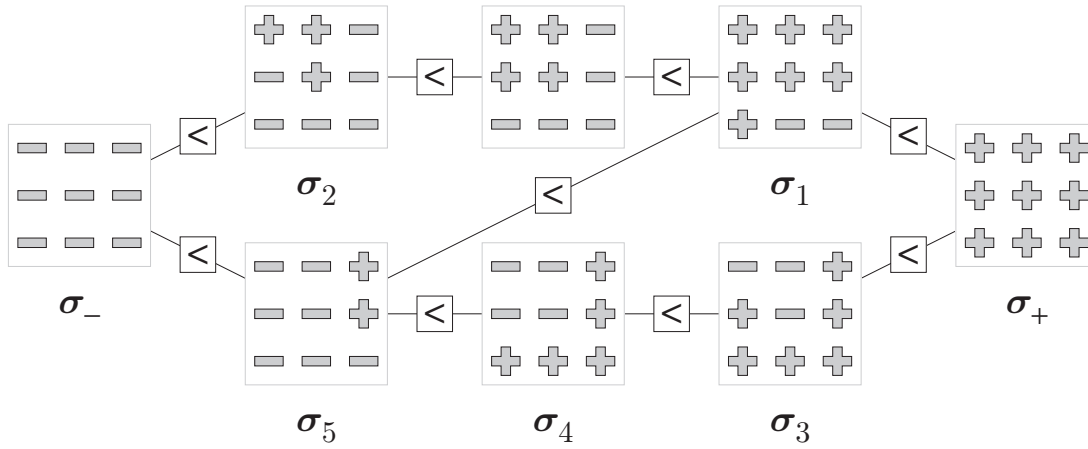


Fig. 5: Partial order in a two-dimensional 3×3 Ising spin model. Order relations are indicated, but not all configurations are related to each other. The all + configuration σ_+ is larger, and the all - configuration σ_- is smaller than all other configurations. The partial order is preserved by the heatbath algorithm (figure taken from [1]).

and it is the advent of new algorithms that clearly make that point (as in [2, 9, 10]). Beginners often think it is sufficient to start the simulation with some “random” initial condition, but this is a misconception, as it is normally impossible to sample an initial configuration from a distribution π^0 anywhere close to π . In the Ising model, a time-honored practice has consisted in starting *two* initial configurations, one from the all up configuration $\sigma_k = +1 \forall k$ and another one from an all down configuration $\sigma_k = -1 \forall k$. According to popular wisdom, when both simulations reached similar magnetizations, they reached equilibrium. This idea turned out to be basically correct. To explain why, we first have to understand the concept of partial order: One configuration a is “smaller” than another one, b , if $\sigma_i^a \leq \sigma_i^b \forall i$. Not all configurations can be compared to each other, but all can be compared to the all up and the all down configuration (see Fig. 5). If we use the same random numbers for choosing the site i and for updating the spin σ_i (coupling), and update the configurations using the heatbath algorithm of Eq. (11), then the partial order is invariant under the dynamics. This means that the children of the all +1 configuration will always stay above the descendants of the all -1 configuration. As soon as both match, at the coupling time τ_{coup} , all possible initial configurations will have produced the same output. This means that the memory of the initial configuration has been lost completely, with the implication that the coupling time is a rigorous upper bound for the convergence time. Propp and Wilson, in a classic paper [12], succeeded in turning the coupling approach into a rigorous exact sampling method (from the distribution π , without *any* corrections). This has far-reaching consequences, most of which have not been explored so far. See [1], Section 1.1.7 and [13] for a detailed discussion of the coupling approach, and of its most surprising finding, namely that a Markov chain can be used to sample π^∞ , without any correction whatsoever.

References

- [1] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, 2006)
- [2] E.P. Bernard, W. Krauth, and D.B. Wilson, Phys. Rev. E **80**, 056704 (2009)
- [3] S.C. Kapfer and W. Krauth, arXiv:1301.4901 (2013)
- [4] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz, J. Comput. Phys. **17**, 10 (1975)
- [5] E.A.J.F. Peters and G. de With, Phys. Rev. E **85**, 026703 (2012)
- [6] M. Michel, S.C. Kapfer, W. Krauth, Manuscript in preparation (2013)
- [7] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989)
- [8] R.H. Swendsen, J.S. Wang, Phys. Rev. Lett. **58**, 86 (1987)
- [9] E.P. Bernard, W. Krauth, Phys. Rev. Lett. **107**, 155704 (2011)
- [10] M. Engel, J.A. Anderson, S.C. Glotzer, M. Isobe, E.P. Bernard, W. Krauth, Phys. Rev. E **87**, 042134 (2013)
- [11] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953)
- [12] J.G. Propp and D.B. Wilson, Rand. Struct. and Alg. **9**, 223 (1996)
- [13] C. Chanal, W. Krauth, Phys. Rev. E **81**, 016705 (2010)

7 Monte Carlo Simulations of Quantum Spin Models

Stefan Wessel

Institute for Theoretical Solid State Physics

RWTH Aachen University

Contents

1	Introduction	2
2	World lines and local updates	2
2.1	Suzuki-Trotter decomposition	3
2.2	World lines	6
2.3	Local updates	7
2.4	The continuous-time limit	8
3	Stochastic series expansion	10
3.1	Configuration space	11
3.2	Observables	13
3.3	Local updates	15
4	The loop algorithm	16
5	Worms, operator loops, and directed loops	21
6	The sign problem	23
A	Quantum Heisenberg Model 101	26

1 Introduction

These lecture notes aim to provide a basic introduction to the central concepts behind modern quantum Monte Carlo (QMC) techniques for the simulation of quantum spin systems in thermal equilibrium. Acquaintance with the Monte Carlo approach to simulate classical statistical physics models is assumed, and can be obtained for example from the lecture of W. Krauth in this volume. In devising QMC simulation methods, two major steps need to be taken: (i) The quantum partition function must be reformulated in a way as to allow stochastic sampling over a space of effective configurations, such that each contributing configuration has a positive statistical weight. (ii) An efficient update-scheme needs to be devised in order to sample this configuration space through a Markov process. The first part of these lecture notes (Secs. 2 and 3) concerns the derivation of different representations of the quantum partition function, and thus relates to the first of these steps. Similarly as in the case of classical Monte Carlo schemes, a major breakthrough concerning the second step was the invention of efficient global update schemes, which will be introduced in the second part of these notes (Secs. 4 and 5). Finally, in section 6, we will discuss the QMC sign problem, which still poses the most severe restriction in applying the presented QMC methods to (frustrated) quantum spin systems. There exist already various excellent recent reviews on the methods that we discuss below, and we will mention some of these at the appropriate places. Thus, in these notes, we do not attempt to provide a comprehensive account on the subject, but instead we will highlight the key ideas behind these techniques.

2 World lines and local updates

We will discuss QMC methods for the simulation of (finite) quantum spin systems at finite temperatures. Upon tuning the simulation temperature sufficiently low (sufficiently below the finite-size spin gap), however, ground state properties of a finite system can usually be explored as well. As will be seen below, these QMC methods are feasible for non-frustrated spin models. In particular, this restricts for example the simulation of the antiferromagnetic Heisenberg model to bipartite lattices. But, there are no restrictions regarding e.g., the dimensionality of the system, and also very large systems with several thousands or even millions of spins can be treated, depending on the temperature range of interest; the computational effort of these methods scales linearly in both the system size and in the inverse temperature $\beta = 1/T$ (we fix $k_B = 1$ here). For the sake of clarity, we consider in the following the spin-1/2 Heisenberg model on an open chain with N_s lattice sites,

$$H = J \sum_{i=1}^{N_s-1} \vec{S}_i \cdot \vec{S}_{i+1} \quad (1)$$

as an example system when deriving the QMC algorithms. We refer to appendix A concerning the notation employed here. In the following, we are interested in calculating thermal expectation values, such as for example the internal energy $E = \langle H \rangle$ of the system at an inverse

temperature β . From statistical physics, we know that

$$\langle H \rangle = \frac{1}{Z} \text{Tr} (H e^{-\beta H}) = \frac{1}{Z} \sum_n E_n e^{-\beta E_n}, \quad (2)$$

where in the last step we expressed the trace in the basis of eigenstates of the Hamiltonian, and the partition function $Z = \text{Tr}(e^{-\beta H}) = \sum_n e^{-\beta E_n}$. While the above expressions in the basis of eigenstates of the Hamiltonian closely resemble the corresponding formulas for a classical system, the problem in the quantum case is that usually one does not know the full spectrum of the Hamiltonian, and thus the above sums cannot be directly calculated. Knowing the spectrum of H in the quantum case, one would often have essentially already solved the problem of interest.

QMC methods circumvent the full diagonalization of H by mapping the quantum partition function Z to a partition function of an effective classical model, and then performing a Monte Carlo sampling of the states contributing to the effective classical partition function. In order to see how such a mapping can be realized, we consider first the Suzuki-Trotter decomposition of the partition function, as pioneered by Suzuki in 1976 [1]. This approach will set the stage for more advanced QMC approaches.

2.1 Suzuki-Trotter decomposition

Our first task will be to map the quantum partition function to that of an effective classical model. In the following, we consider in particular the quantum Heisenberg model on an open chain, given by the Hamiltonian Eq. (1), which can be decomposed into contributions from each bond,

$$H = \sum_i H_i, \quad H_i = J \vec{S}_i \cdot \vec{S}_{i+1}, \quad (3)$$

where H_i is a bond Hamiltonian, that corresponds to the bond between site i and site $i + 1$ on the chain. We can furthermore separate H into two parts,

$$H = H_A + H_B, \quad (4)$$

where H_A (H_B) consists of only the even (odd) bond contributions, i.e.,

$$H_A = \sum_{i \text{ even}} H_i, \quad H_B = \sum_{i \text{ odd}} H_i. \quad (5)$$

While $[H_A, H_B] \neq 0$, the bond Hamiltonians within H_A or H_B commute among themselves, since $[H_i, H_j] = 0$ for i and j both even (or both odd). We have thus separated H into two parts, each of which consists of commuting terms. Now, we rewrite the statistical operator as a product of many terms, each with a small prefactor β/M in front of H ,

$$e^{-\beta H} = \left(e^{-\frac{\beta}{M} H} \right)^M = \left(e^{-\Delta\tau H} \right)^M, \quad (6)$$

with a (large) integer number (called the Trotter number) M , and $\Delta\tau = \beta/M$. The Suzuki-Trotter (or split-operator) approximation [1, 2] now consists in approximating the exponential

of H , expressed in terms of the two non-commuting pieces $H = H_A + H_B$, by a product of exponentials. While many such decompositions are possible, the most commonly used approximations (also employed beyond QMC methods) are

$$e^{-\Delta\tau H} = e^{-\Delta\tau(H_A+H_B)} = \begin{cases} e^{-\Delta\tau H_A} e^{-\Delta\tau H_B} & +O(\Delta\tau^2) \\ e^{-\Delta\tau H_B/2} e^{-\Delta\tau H_A} e^{-\Delta\tau H_B/2} & +O(\Delta\tau^3) \end{cases}, \quad (7)$$

where the errors are also proportional to the commutator $[H_A, H_B]$. Using either of the two approximations in Eq. (6), we obtain

$$\begin{aligned} Z &= \text{Tr} (e^{-\beta H}) \\ &= \text{Tr} \left(\underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_M \underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_{M-1} \dots \underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_1 \right) + O(\Delta\tau^2). \end{aligned} \quad (8)$$

When using the second-order approximation from Eq. (7), the final expression is obvious. However, when using the third-order approximation, one actually obtains the same final expression, due to the cyclic invariance of the trace, which can be used to move the most left exponential to the very right, and then coalesce every other two consecutive terms. This also shows, that the systematic error in Z actually scales as $M\Delta\tau^3 \propto \Delta\tau^2$, even when using the second-order approximation, where one might have expected an $O(\Delta\tau)$ error in the final expression. Now, consider a basis of the Hilbert space, e.g., in terms of the local eigenstates of S_i^z , which we write

$$\begin{aligned} |\sigma_1\rangle &= |\uparrow\uparrow \dots \uparrow\uparrow\rangle \\ |\sigma_2\rangle &= |\uparrow\uparrow \dots \uparrow\downarrow\rangle \\ &\vdots \\ |\sigma_{2^N}\rangle &= |\downarrow\downarrow \dots \downarrow\downarrow\rangle. \end{aligned} \quad (9)$$

Since this set forms a basis, we obtain a completeness relation:

$$\sum_{\sigma} |\sigma\rangle\langle\sigma| = 1, \quad (10)$$

where $|\sigma\rangle\langle\sigma|$ is a projection operator onto the basis state $|\sigma\rangle$. Within this basis of the Hilbert space, we can thus express the partition function as

$$\begin{aligned} Z &= \sum_{\sigma^0} \langle\sigma^0| e^{-\beta H} |\sigma^0\rangle \\ &\approx \sum_{\sigma^0} \langle\sigma^0| \underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_M \underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_{M-1} \dots \underbrace{e^{-\Delta\tau H_A} e^{-\Delta\tau H_B}}_1 |\sigma^0\rangle \\ &= \sum_{\sigma^0} \langle\sigma^0| e^{-\Delta\tau H_A} e^{-\Delta\tau H_B} \dots e^{-\Delta\tau H_A} \left(\sum_{\sigma^1} |\sigma^1\rangle\langle\sigma^1| \right) e^{-\Delta\tau H_B} |\sigma^0\rangle \\ &= \sum_{\sigma^0} \sum_{\sigma^1} \langle\sigma^0| e^{-\Delta\tau H_A} e^{-\Delta\tau H_B} \dots \left(\sum_{\sigma^2} |\sigma^2\rangle\langle\sigma^2| \right) e^{-\Delta\tau H_A} |\sigma^1\rangle\langle\sigma^1| e^{-\Delta\tau H_B} |\sigma^0\rangle \\ &= \sum_{\sigma^0, \sigma^1, \sigma^2} \langle\sigma^0| e^{-\Delta\tau H_A} e^{-\Delta\tau H_B} \dots \left(\sum_{\sigma^3} |\sigma^3\rangle\langle\sigma^3| \right) e^{-\Delta\tau H_B} |\sigma^2\rangle\langle\sigma^2| e^{-\Delta\tau H_A} |\sigma^1\rangle\langle\sigma^1| e^{-\Delta\tau H_B} |\sigma^0\rangle. \end{aligned}$$

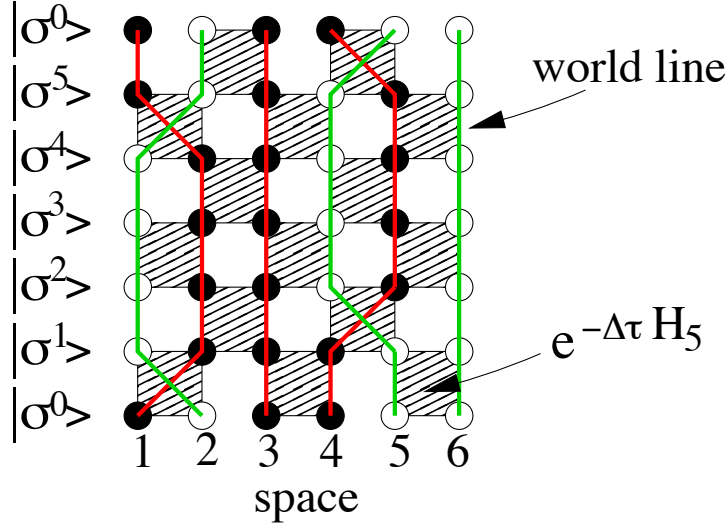


Fig. 1: A configuration contributing to the checkerboard decomposition of the partition function of an open spin-1/2 Heisenberg chain with 6 lattice sites, for a Trotter number $M = 3$. Black (white) circles denote spin up (down), and the world lines of the up (down) spins are denoted by red (green) lines. Each dashed square represents the exponential of a bond Hamiltonian.

In the above summations, we added a (superscript) label to distinguish the basis sets that arise from the trace ($|\sigma^0\rangle$) and from the various inserted partitions of unity ($|\sigma^1\rangle$, $|\sigma^2\rangle$, and $|\sigma^3\rangle$). Continuing this way, we eventually arrive at the following expression:

$$Z = \sum_{\{\sigma^i\}} \underbrace{\langle \sigma^0 | e^{-\Delta\tau H_A} | \sigma^{2M-1} \rangle}_{2M} \underbrace{\langle \sigma^{2M-1} | e^{-\Delta\tau H_B} | \sigma^{2M-2} \rangle}_{2M-1} \underbrace{\langle \sigma^{2M-2} | e^{-\Delta\tau H_A} | \sigma^{2M-3} \rangle}_{2M-2} \dots \underbrace{\langle \sigma^3 | e^{-\Delta\tau H_B} | \sigma^2 \rangle}_3 \underbrace{\langle \sigma^2 | e^{-\Delta\tau H_A} | \sigma^1 \rangle}_2 \underbrace{\langle \sigma^1 | e^{-\Delta\tau H_B} | \sigma^0 \rangle}_1 + \mathcal{O}(\Delta\tau^2). \quad (11)$$

This representation of Z is referred to as the “Suzuki-Trotter decomposition”. Furthermore, since H_A and H_B each consists of commuting parts, we find that

$$e^{-\Delta\tau H_A} = \prod_{i \text{ even}} e^{-\Delta\tau H_i} = e^{-\Delta\tau H_2} e^{-\Delta\tau H_4} \dots \quad (12)$$

$$e^{-\Delta\tau H_B} = \prod_{i \text{ odd}} e^{-\Delta\tau H_i} = e^{-\Delta\tau H_1} e^{-\Delta\tau H_3} \dots \quad (13)$$

and thus in the Suzuki-Trotter decomposition each exponential expression factorizes into exponentials for the bond Hamiltonians on either all even or all odd bonds. One can represent a given contribution to Z graphically, as is shown for a specific example with $N_s = 6$, and $M = 3$ in Fig. 1. This two-dimensional picture appears like a space-time picture of spins propagating in discrete steps from the initial configuration $|\sigma^0\rangle$ to $|\sigma^1\rangle, \dots$, and finally from $|\sigma^{2M-1}\rangle$ back to $|\sigma^0\rangle$, since the configurations on the first and the last step are equal. This pictorial illustration also explains why the underlying partitioning of the Hamiltonian is often referred to as the “checkerboard decomposition”.

2.2 World lines

Each exponential $e^{-\Delta\tau H_i}$ couples two spin sites on a shaded plaquette, and the statistical weight of a configuration is the product of all the matrix elements of the exponentials on all shaded plaquettes. Note, that $e^{-\Delta\tau H_i}$ appears as an imaginary-time evolution operator, that propagates the two spins at sites i and $i+1$ from one “time slice” to the next. To appreciate this suggestive picture of a discrete time propagation, we introduced the corresponding “Trotter time-step” $\Delta\tau = \beta/M$ above. However, not all of the $2^4 = 16$ possible spin configurations along a shaded plaquette are allowed to appear within the checkerboard decomposition. In order to assess the allowed configurations and the corresponding matrix elements, we explicitly calculate the matrix elements of the operators

$$e^{-\Delta\tau H_i} = \sum_{k=0}^{\infty} \frac{1}{k!} (\Delta\tau)^k (-H_i)^k. \quad (14)$$

In the local two-sites basis of the two spins at site i and $i+1$, the Hamiltonian matrix is

$$H_i = J \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & -\frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}, \quad \text{in the local basis} \quad \begin{matrix} |\uparrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{matrix}. \quad (15)$$

Upon performing the Taylor expansion, and grouping back the resulting terms, one finds (this would make up a nice exercise)

$$e^{-\Delta\tau H_i} = e^{+\Delta\tau J/4} \begin{pmatrix} e^{-\frac{\Delta\tau J}{2}} & 0 & 0 & 0 \\ 0 & \cosh(\frac{\Delta\tau J}{2}) & -\sinh(\frac{\Delta\tau J}{2}) & 0 \\ 0 & -\sinh(\frac{\Delta\tau J}{2}) & \cosh(\frac{\Delta\tau J}{2}) & 0 \\ 0 & 0 & 0 & e^{-\frac{\Delta\tau J}{2}} \end{pmatrix}. \quad (16)$$

There are thus 6 possible allowed plaquettes with a finite weight. They are shown in Fig. 2. All other combinations would lead to a vanishing matrix element and are thus not allowed to occur as part of any allowed configuration. The allowed configurations are those that exhibit the same total local magnetization $S_i^z + S_{i+1}^z$ on the lower and the upper edge of each shaded plaquette. This property derives from the fact that each bond Hamiltonian H_i conserves the total magnetization of the two spins connected by the bond.

When we connect the positions of the up and down spins as they propagate through the shaded plaquettes, we obtain continuous lines. These are the “world lines” of the spins, and each such world line denotes the evolution of one spin up or down from $|\sigma^0\rangle$ back to $|\sigma^0\rangle$. The world lines for both the spin-up and the spin-down case are illustrated in Fig. 1. Note, that it would be sufficient to only show, say, the up spins, which is what will be done later on in these notes. If one would glue together the upper and lower boundary of the space-time configuration to make the periodicity constraint by the trace more explicit, these world lines are thus continuous, and are not broken anywhere. This set of unbroken world-line configurations defines a classical



Fig. 2: Allowed shaded plaquette configurations for the spin-1/2 Heisenberg model. Black (white) circles denote spin up (down), and the world lines of the up (down) spins are denoted by red (green) lines.

statistical model, which has the same partition function as the quantum partition function, if the weight of a given world-line configuration C equals the product of the matrix elements from all shaded plaquettes P

$$W(C) = \prod_P W_P(C|_P), \quad (17)$$

where $W_P(C|_P)$ is the corresponding matrix element of $e^{-\Delta\tau H_i}$ for plaquette P in the configuration C . At first sight, the effective model might appear to be just the two-dimensional Ising model on a square lattice of size $N_s \times 2M$, with periodic boundary conditions in the y -direction. However, the effective classical model is in fact more complex than the Ising model, which has an unconstrained configuration space. Only a subset of configurations of the two-dimensional Ising model are allowed to occur also in the effective classical model for Z . These configurations are those that correspond to continuous, unbroken world lines. In higher spatial dimensions, one can use the Suzuki-Trotter approach with decompositions very similar to the one discussed above. One thus finds that the quantum partition function of a d -dimensional system (here $d = 1$) corresponds to that of an effective classical model in $(d + 1)$ dimensions, with the $(d + 1)$ -th direction corresponding to an imaginary-time evolution of the original quantum model. This well known quantum-to-classical-mapping holds in fact much more generally, and exposes a deep connection between classical and quantum statistical physics.

Note however the minus sign in front of the off-diagonal matrix elements: for the allowed world-line configurations, all the accumulated signs actually cancel out, due to the periodicity constraint in imaginary time, in case of the open chain or for a closed chain with an even number of sites. This holds also true for higher-dimensional generalizations of the world-line approach whenever the underlying lattice structure is bipartite. However, for frustrated systems, e.g., for a closed 3-site chain, there appears a “sign problem”, since both configuration with positive and negative weights are allowed. We will discuss the consequences of this issue in Sec. 6.

2.3 Local updates

The world-line representation introduced above can be taken as a starting point to set up a QMC algorithm [3]. But, how does one sample configurations in the effective world-line model in a Monte Carlo algorithm, i.e., how do we generate new valid world-line configurations from a given one? Since the spin S^z -conservation prohibits the breaking of world lines, the updates need to move world-line segments instead of just changing local spin states like in the classical Ising model. In the early days, so-called local updates were performed, i.e., local manipulations

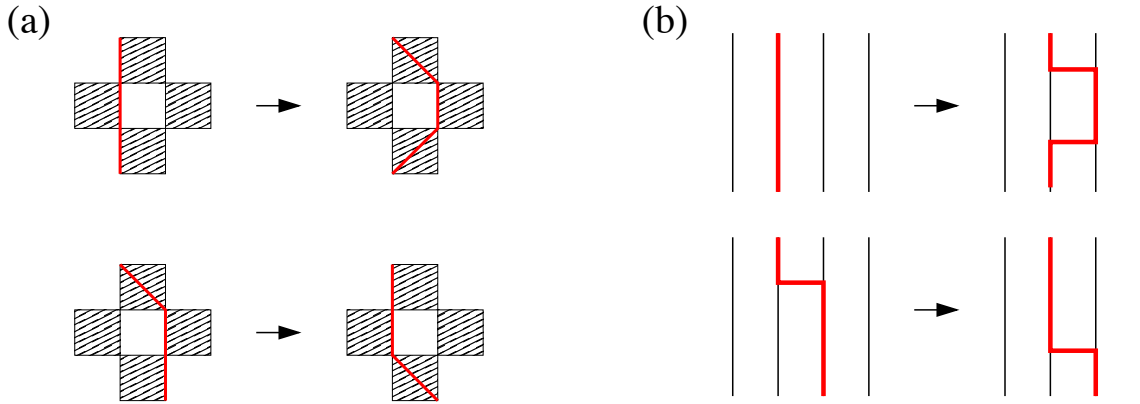


Fig. 3: Two different local updates of world-line configurations in discrete time (a) and in the continuous-time limit (b). For clarity, only the spin-up world lines are shown explicitly.

of the world-line configuration, such as those shown in Fig. 3 (a). These local updates are quite simple and either generate a pair of kinks across a white plaquette within a straight section of the world line, or they move a kink to a neighboring plaquette [3, 4], as shown in Fig. 3 (a). Slightly more complicated local moves are required for higher-dimensional spin systems [5], or for other quantum lattice models [6]. In any case, such local updates are accepted by considering the related change in the statistical weight according to Eq. (17), using e.g., the Metropolis [7] acceptance scheme.

However, such local updates are not efficient in two respects: First, global properties of the configurations cannot be changed by such local updates; the spatial winding number of the world lines and their total number, i.e. the total magnetization of the system, cannot be changed using local moves. Thus, they had to be complemented by special global updates as well [5]. Second, such updates lead to severe critical slowing down upon approaching critical regions, similarly as those familiar from the local Metropolis algorithm for the Ising model. A major breakthrough in overcoming this problem was the work by Evertz, Lana and Marcu [8], in which they presented an extension of the Swendsen-Wang cluster update idea known from classical Monte Carlo studies to world-line QMC methods [56]. We will present this “loop algorithm” in Sec. 4. Before doing so, we want to introduce two other QMC representations, which are relevant for modern QMC algorithms.

2.4 The continuous-time limit

Due to the finite number of time slices, M , the approach described above suffers from a systematic error – the discretization or Trotter error. It has been shown that in most cases, one can keep $\Delta\tau$ independent of N_s and β in order to ensure a constant error level [9] (for some observables, however, care has to be taken to avoid divergent errors in the zero temperature limit $\beta \rightarrow \infty$ [9, 10]). The Trotter error was controlled originally by extrapolation to the continuous-time limit $\Delta\tau \rightarrow 0$, from simulations with different values of the time step $\Delta\tau$. It was realized later [11] that the continuous-time limit can be taken already in the construction of

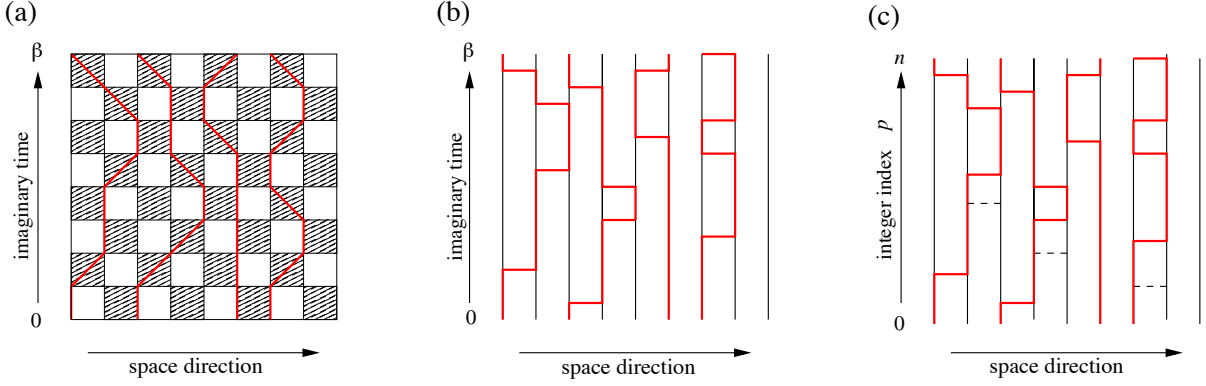


Fig. 4: Comparison of world-line QMC configurations in a) discrete time, b) continuous time, and c) in SSE representation. In the latter, the continuous-time index is replaced by an ordered integer index of the operators, and additional diagonal terms are indicated by dashed lines.

the algorithm, so that simulations can be performed directly in the limit $\Delta\tau \rightarrow 0$, i.e., $M \rightarrow \infty$, without the need to perform any final extrapolations. This appears feasible, once one realizes that a given world-line configuration can be represented by keeping only a list of times at which the configuration *changes*, instead of storing the configuration at each of the $2M$ time slices. Indeed, the probability for a jump of a world line (i.e., a kink in the world line) from one lattice site to a neighboring site across a given plaquette, and thus a change in the local configuration is proportional to $\sinh(\Delta\tau J/2) \propto \Delta\tau \propto 1/M$. Hence, the mean total number of such kinks remains finite in the limit $M \rightarrow \infty$. The continuous world-line representation, i.e., the limit $\Delta\tau \rightarrow 0$, is thus well defined, and the relevant configurational information can be efficiently represented on a computer. This is illustrated in Fig. 4 (a) and (b), by comparing the discrete and continuous-time world-line approach.

In addition to the configurational information, the local updates need to be revisited in the continuous-time limit as well. In particular, the probability P_{pk} for the insertion of a pair of kinks in the world line (the upper move in Fig. 3 (a)), vanishes in the continuous-time limit as

$$P_{pk} = \sinh^2(\Delta\tau J/2) / \cosh^2(\Delta\tau J/2) \propto \Delta\tau^2 \propto 1/M^2 \rightarrow 0. \quad (18)$$

To circumvent such a vanishing probability, one now proposes to insert a pair of jumps not at specific locations, but *anywhere* inside a finite time interval [11], as illustrated in Fig. 3 (b). The integrated probability for such a move then remains finite in the continuous-time limit. Similarly, instead of shifting a kink by $\Delta\tau$ (the lower move in Fig. 3 (a)), one now moves it within a finite time interval in the continuous-time algorithm. In addition to local updates, also the loop algorithm has been shown to allow for an efficient realization in the continuous-time limit [12], as will be discussed below.

Finally, we remark that the continuous-time limit of the Suzuki-Trotter formula, Eq. (11), in the above interpretation is in fact equivalent to a time-dependent perturbation theory in imaginary

time,

$$\begin{aligned} Z &= \text{Tr} \exp(-\beta H) = \text{Tr} \left[\exp(-\beta H_0) \mathcal{T} \exp \int_0^\beta d\tau V(\tau) \right] \\ &= \text{Tr} \left[\exp(-\beta H_0) \left(1 - \int_0^\beta d\tau V(\tau) + \frac{1}{2} \int_0^\beta d\tau_1 \int_{\tau_1}^\beta d\tau_2 V(\tau_1) V(\tau_2) + \dots \right) \right], \end{aligned} \quad (19)$$

where the symbol \mathcal{T} denotes the (imaginary-) time ordering of the exponential [11, 13]. In this representation, the Hamiltonian $H = H_0 + V$ is split up into a diagonal term H_0 and an off-diagonal perturbation V , which in the interaction picture is $V(\tau) = \exp(\tau H_0) V \exp(-\tau H_0)$. For the Heisenberg model, the diagonal term H_0 is given by the longitudinal $S_i^z S_j^z$ spin-spin interactions, while the off-diagonal perturbation V relates to the transverse, spin exchange terms, $\frac{1}{2}(S_i^+ S_j^- + S_i^- S_j^+)$ in H . In more detail, we express the Heisenberg Hamiltonian as

$$H = H_0 + V = \sum_{\langle i,j \rangle} J S_i^z S_j^z + \sum_{\langle i,j \rangle} \frac{J}{2} (S_i^+ S_j^- + S_i^- S_j^+). \quad (20)$$

The number of kinks in a given continuous-time world-line configuration is then equal to the expansion order of a specific term in the perturbation expansion, wherein each kink's space-time position is set by a specific spin exchange term from V between two lattice sites i and j , as well as a corresponding imaginary time τ between 0 and β . Furthermore, the exponential factors $\exp(\tau H_0)$ in $V(\tau)$ describe the vertical, unperturbed evolution of the spins along the world lines between the kinks, thus establishing explicitly the above-mentioned equivalence between these two formulations of the continuous-time limit.

3 Stochastic series expansion

Already before Suzuki's approach to QMC in the mid 1970s [1, 3], an alternative approach to QMC simulations had been put forward by Handscomb in the early 1960s [14, 15] for the specific case of the ferromagnetic Heisenberg model. It is based on a Taylor expansion of the statistical operator inside the partition function:

$$Z = \text{Tr} \exp(-\beta H) = \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \text{Tr} ((-H)^n). \quad (21)$$

In Handscomb's approach and within later extensions to other models [16–18], the traces of H^n were evaluated employing projection operator expressions. Later, in the early 1990s, the power-series approach to QMC simulations was revisited by Sandvik and Kurkijärvi within the more generally applicable stochastic series expansion (SSE) formulation [19, 20], in which these traces are also sampled stochastically. A recent review of the SSE method, combined with a general introduction to computational methods for quantum spin systems, can be found in Ref. [21]. It contains also many basic details for implementing the SSE QMC method.

3.1 Configuration space

To formulate the SSE method, it proves convenient to first express the Heisenberg Hamiltonian as a sum of bond operators, which are either diagonal or off-diagonal in the standard S^z -basis,

$$H = - \sum_{b=1}^{N_b} (H_{1,b} - H_{2,b}) + JN_b/4, \quad (22)$$

with

$$H_{1,b} = J \left(\frac{1}{4} - S_{i(b)}^z S_{j(b)}^z \right), \quad (23)$$

$$H_{2,b} = \frac{J}{2} \left(S_{i(b)}^+ S_{j(b)}^- + S_{i(b)}^- S_{j(b)}^+ \right). \quad (24)$$

Here, N_b equals the total number of bonds in the system, and $i(b)$ and $j(b)$ denote the two lattice sites that are connected by the bond with bond index b . The explicit introduction of the minus sign in H is convenient, while the constant in the diagonal operators will make the series expansion positive definite, as we will see shortly. Note that this constant is irrelevant for the physics of the system, but has to be accounted for when calculating the system's energy, as we will see in the next section. We now insert the above form of the Hamiltonian into the expression Eq. (21) for the partition function, thereby obtaining

$$Z = \sum_{\sigma} \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \sum_{\mathcal{S}_n} (-1)^{n_2} \langle \sigma | H_{t(n),b(n)} \cdots H_{t(2),b(2)} H_{t(1),b(1)} | \sigma \rangle. \quad (25)$$

Here we used the local S^z -basis to express the trace, and \mathcal{S}_n denotes products (strings) of the bond operators $H_{1,b}$ or $H_{2,b}$ that originate from expanding H^n in these bond operator terms using Eq. (22), namely, each such operator string \mathcal{S}_n is an sequence of n bond operators, specified by the type labels $t(p) \in \{1, 2\}$ (i.e. diagonal or off-diagonal) and $b(p) \in \{1, \dots, N_b\}$, so that we can also write such a sequence as

$$\mathcal{S}_n = [t(1), b(1)], [t(2), b(2)], \dots, [t(n), b(n)]. \quad (26)$$

The summation in Eq. (25) extends over all expansion orders n , and for each expansion order, over all operator sequences \mathcal{S}_n of length n , containing n bond operators. Finally, n_2 denotes the number of off-diagonal operators in the sequence \mathcal{S}_n , i.e., the number of sequence elements with $t(p) = 2$. It is important to note that, in the local S^z -basis, the bond operators exhibit a non-branching property, i.e., if one applies a bond operator to one of these basis states, the resulting state is either proportional to the same basis state, or to another basis state, or it vanishes. But in no case it is a superposition of two or more such basis states generated when traversing the action of the sequence of bond operators starting from any initial basis state $|\sigma\rangle$.

The propagated basis state that appears after the action of the first p operators, i.e., at propagation level p , will be denoted (after proper normalization) by $|\sigma(p)\rangle$, i.e.,

$$|\sigma(p)\rangle \propto \prod_{q=1}^p H_{t(q),b(q)} |\sigma\rangle. \quad (27)$$

Note that for any operator-state configuration $(\mathcal{S}_n, |\sigma\rangle)$ that contributes to Z , the final state, resulting after the action of all the n operators has to fulfill $\langle\sigma|\sigma(n)\rangle \neq 0$, thus we obtain the following periodicity constraint: $|\sigma(n)\rangle = |\sigma(0)\rangle = |\sigma\rangle$.

If one examines in more detail the action of the different bond operators $H_{1,b}$ and $H_{2,b}$ on the two spins at the related lattice sites $i(b)$ and $j(b)$ of a basis state $|\sigma\rangle$, one finds, that (i) the state is destroyed if both spins are parallel,

$$H_{1,b}|\uparrow_{i(b)}\uparrow_{j(b)}\rangle = 0, \quad H_{1,b}|\downarrow_{i(b)}\downarrow_{j(b)}\rangle = 0, \quad (28)$$

$$H_{2,b}|\uparrow_{i(b)}\uparrow_{j(b)}\rangle = 0, \quad H_{2,b}|\downarrow_{i(b)}\downarrow_{j(b)}\rangle = 0, \quad (29)$$

and (ii) if the two spins are anti-parallel, the spin state is preserved by $H_{1,b}$, while being flipped by $H_{2,b}$, in both cases with a matrix element of $J/2$,

$$\langle\uparrow_{i(b)}\downarrow_{j(b)}|H_{1,b}|\uparrow_{i(b)}\downarrow_{i(b)}\rangle = J/2, \quad \langle\downarrow_{i(b)}\uparrow_{j(b)}|H_{1,b}|\downarrow_{i(b)}\uparrow_{i(b)}\rangle = J/2, \quad (30)$$

$$\langle\downarrow_{i(b)}\uparrow_{j(b)}|H_{2,b}|\uparrow_{i(b)}\downarrow_{i(b)}\rangle = J/2, \quad \langle\uparrow_{i(b)}\downarrow_{j(b)}|H_{2,b}|\downarrow_{i(b)}\uparrow_{i(b)}\rangle = J/2. \quad (31)$$

From (i) it follows that for any operator-state configuration $(\mathcal{S}_n, |\sigma\rangle)$ that contributes to Z , the bond operators $H_{t(p),b(p)}$ act only on propagated states with anti-parallel spins on bond $b(p)$. Property (ii) implies that all the matrix elements that arise from the bond operators within the operator string are equal and positive. The effective configuration space of the SSE method thus essentially consists of all allowed operator-state configurations, i.e., those with non-zero weight, with the weight of a given configuration $C = (\mathcal{S}_n, |\sigma\rangle)$ obtained from Eq. (25),

$$Z = \sum_C W(C), \quad W(C) = \frac{\beta^n}{n!} \langle\sigma| \prod_{p=1}^n H_{t(p),b(p)} |\sigma\rangle. \quad (32)$$

Here, we also anticipated that on a bipartite lattice the number of off-diagonal operators in an allowed operator sequence must be even, such that $(-1)^{n_2} = +1$, due to the periodicity constraint $|\sigma(n)\rangle = |\sigma(0)\rangle = |\sigma\rangle$. For a non-bipartite (frustrated) lattice this would not be the case, and a QMC sign-problem would result; cf. Sec. 6 for a discussion of this case. The SSE QMC configurations can be visualized very similarly as in the previous case, cf. Fig. 4 for a comparison between the different formulations. Note that in the SSE formulation a discrete time-like index p is introduced. Furthermore, in addition to the jump events that appear in the continuous-time world-line formulation, the SSE configurations exhibit the additional presence of diagonal bond-operators. Indeed, the SSE representation can be formally related to the continuous-time world-line representation by observing that Eq. (25) is obtained from Eq. (19) upon setting $H_0 = 0$, $V = H$ and integrating over all times (compare also Fig. 4) [22]. This mapping exposes the respective advantages and disadvantages of the two representations: The SSE representation corresponds to a perturbation expansion in all terms of the Hamiltonian, whereas continuous-time world-line algorithms treat the diagonal terms exactly and perturb only in the off-diagonal terms of the Hamiltonian. The continuous-time algorithm hence needs fewer terms in the expansion, but we pay for it by having to handle continuous imaginary-time variables.

For spin systems in which the contributions to the total energy arising from diagonal and off-diagonal terms are well balanced, the SSE is typically preferable in practice, due to its discrete representation of the imaginary-time continuum. It is important to note, that this discretization does *not* introduce any Trotter-error as in the discrete-time world-line QMC; instead, the SSE formulation should be considered an essentially exact QMC technique.

It is possible to formulate Monte Carlo sampling algorithms for the SSE configuration space introduced above, i.e., containing operator strings of fluctuating length [14, 20, 23]. However, it is also possible to work with operator strings of a fixed length, which is in most cases more convenient computationally [19, 20]. Indeed, as will be shown in the next section, the mean length of the operator string $\langle n \rangle$ is essentially related to the energy of the system and scales linear with the system size N_s and β . Furthermore, the specific heat is given by $C = \langle n^2 \rangle - \langle n \rangle^2 - \langle n \rangle$, so that in particular for $T \rightarrow 0$, where C vanishes, the variance of the distribution of expansion orders is seen to be equal to $\langle n \rangle$. Hence, for a given system size and temperature, the series expansion order n of the operator-state configurations $C = (\mathcal{S}_n, |\sigma\rangle)$ can in practice be truncated at a sufficiently large cutoff Λ , which is typically determined during the thermalization phase of an SSE simulation, without introducing any detectable error. It is then feasible to work with fixed-length operator strings upon augmenting all operator sequences of length $n < \Lambda$ by $\Lambda - n$ unit operators I , denoted in the following by $H_{0,0} = I$. Allowing for all possible positions of the unit operators in the original operator strings, one is then lead to the following representation of the partition function:

$$Z = \sum_C W_\Lambda(C), \quad W(C) = \frac{\beta^n (\Lambda - n)!}{\Lambda!} \langle \sigma | \prod_{p=1}^{\Lambda} H_{t(p), b(p)} | \sigma \rangle, \quad (33)$$

where each operator-state configuration $(\mathcal{S}_\Lambda, |\sigma\rangle)$ now consists of a sequence of operators specified by the type labels $t(p) \in \{0, 1, 2\}$ and $b(p) \in \{1, \dots, N_b\}$ in case $t(p) \in \{1, 2\}$, or $b(p) = 0$, if $t(p) = 0$, which would represent a unit operator. In the above formula, n denotes the number of non-unit operators in the operator sequence, i.e., the number of true bond operators from H . The new combinatorial factor in $W_\Lambda(C)$ accounts for the $\binom{\Lambda}{n}$ equivalent terms that are generated this way as compared to the expression without unit operators in Eq. (25). Again, we left out the factor $(-1)^{n_2} = +1$ on bipartite lattices, where all allowed configurations thus contribute with a positive weight $W_\Lambda(C) > 0$ to the partition function, which is hence feasible for Monte Carlo sampling.

3.2 Observables

Thus far in this exposition of QMC methods, we focused on different representations of the quantum partition function in terms of an effective set of configurations that prove amenable to Monte Carlo sampling. However, we did not mention yet how observables can be calculated. Here, we provide an overview how typical observables are measured within the SSE framework [19, 20, 24]. In many cases, the generalization from the formulas provided below to the other world-line methods is straightforward. A remarkably simple formula holds for the total

inner energy $E = \langle H \rangle$, which can be expressed in terms of the mean expansion order:

$$\langle H \rangle = \frac{1}{Z} \text{Tr}[H e^{-\beta H}] = \frac{1}{Z} \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \text{Tr}[(-H)^{n+1}] = -\frac{1}{Z} \sum_{n=1}^{\infty} \frac{\beta^n}{n!} \frac{n}{\beta} \text{Tr}[(-H)^n] = -\frac{\langle n \rangle}{\beta}. \quad (34)$$

Accounting for the constant shift that we added to the Hamiltonian in Eq. (22) to perform the SSE expansion, we thus obtain the estimator for the inner energy before the shift as

$$E = -\frac{\langle n \rangle}{\beta} + JN_b/4. \quad (35)$$

By the same procedure, that leads to Eq. (34), one finds that $\langle H^2 \rangle = \langle n(n-1) \rangle / \beta^2$, from which the specific heat estimator

$$C = \beta^2 (\langle H^2 \rangle - \langle H \rangle^2) = \langle n^2 \rangle - \langle n \rangle^2 - \langle n \rangle \quad (36)$$

is readily obtained, which we already advocated above. Note, that the constant introduced in Eq. (22) cancels out in the estimator for C , as would be expected. In practice, the statistical error of C becomes large at low temperatures (where C itself becomes small), since it is given by the difference of two large numbers ($\sim (N_s \beta)^2$). We can also readily obtain the expectation value of any operator A that is diagonal in the local S^z -basis, such as for example a spin-spin correlation function $\langle S_i^z S_j^z \rangle$, in which case $A = S_i^z S_j^z$, by using averages over the propagated states:

$$\langle A \rangle = \left\langle \frac{1}{n} \sum_{p=0}^{n-1} A(p) \right\rangle, \quad (37)$$

where $A(p) = \langle \sigma(p) | A | \sigma(p) \rangle$ has been introduced for convenience. From the spin-spin correlation function, the equal-time structure factor is then obtained by Fourier transformation.

Finally, let us mention, how one can measure generalized susceptibilities related to two diagonal operators A, B within the SSE approach. Consider that we add to the Hamiltonian H a linear coupling term $h_B B$, i.e., $H \rightarrow H - h_B B$, then one obtains for the linear response function

$$\chi_{AB} = \left. \frac{\partial \langle A(h_B) \rangle}{\partial h_B} \right|_{h_B=0} \quad (38)$$

the Kubo formula

$$\chi_{AB} = \int_0^\beta d\tau \left(\langle A(\tau) B(0) \rangle - \langle A \rangle \langle B \rangle \right), \quad (39)$$

with $A(\tau) = e^{-\tau H} A e^{\tau H}$ the imaginary time-evolved operator. In case A and B are both diagonal in the local S^z -basis, this integral can be evaluated using

$$\int_0^\beta d\tau \langle A(\tau) B(0) \rangle = \left\langle \frac{\beta}{n(n+1)} \left(\sum_{p=0}^{n-1} A(p) \right) \left(\sum_{p=0}^{n-1} B(p) \right) + \frac{\beta}{(n+1)^2} \left(\sum_{p=0}^n A(p) B(p) \right) \right\rangle. \quad (40)$$

A special case concerns the uniform magnetic susceptibility, for which $A = B = \frac{1}{N_s} \sum_{i=1}^{N_s} S_i^z$, and thus we obtain the simple expression

$$\chi_u = \frac{\beta}{N_s} \langle M_z^2 \rangle, \quad M_z = \sum_{i=1}^{N_s} S_i^z. \quad (41)$$

For further observables, such as the spin stiffness, we refer to Sandvik's review [21]. Let us finally note, that within the world-line approach and the SSE representation it is also possible to evaluate the expectation values of imaginary-time-dependent correlation functions [19, 24]. For an efficient evaluation of the corresponding SSE expressions, based on the explicit mapping between the SSE configuration space and the continuous-time representation, we refer to the review by Assaad and Evertz [25].

3.3 Local updates

In order to sample the SSE configuration space, we need to generate operator-state configurations $(\mathcal{S}_A, |\sigma\rangle)$ according to the appropriate statistical weights. Updates of the operator sequence can in general not be carried out without affecting also the spin configuration. For example, if one changes the operator at a given propagation level p , say from a diagonal operator $H_{1,b}$ on the bond b to the off-diagonal operator $H_{2,b}$ on the same bond, then the spin configuration $|\sigma(p)\rangle$ has to change as well. This change in the world-line configuration must be healed at some other propagation level p' , by also exchanging diagonal and off-diagonal operators, so that the resulting state $|\sigma(p')\rangle$ is the same as before. Hence, one needs to perform Monte Carlo updates by attempting to change the bond operators at two appropriately picked propagation levels. Indeed, in most cases update attempts with randomly picked positions would not be possible, such that one has to specifically search for operator pairs (or even more operators), which can be updated this way. In the end, such a procedures correspond to the local update schemes that we introduced above for the world-line QMC approach, and they suffer from similar problems.

However, more efficient update schemes have been developed [26, 27]. Such updates in fact consist of two sub-steps: (i) the diagonal update, wherein the expansion order n is modified while keeping fixed the spin configuration $|\sigma\rangle$ of $(\mathcal{S}_A, |\sigma\rangle)$ as well as the off-diagonal operator content of the operator sequence \mathcal{S}_A , and (ii) a non-diagonal update, the operator-loop update, that updates the operator content as well as the spin configuration simultaneously – similar to, but in a much more efficient way than, the two-operator case that we discussed above.

We will outline the operator-loop update in the next section, and concentrate here on the diagonal update step, that is an essential local update procedure, which we formulate here for the fixed operator-string length representation [26]. Within this diagonal update, the whole operator sequence is sequentially traversed, and attempts are made to exchange diagonal operators and identity operators, so that in each such step, the expansion order, i.e., the number of non-identity operators, n can change by one to $n \pm 1$. While moving through the operator sequence, the spin configuration $|\sigma\rangle$ is updated whenever an off-diagonal operator is encountered, such that the propagated spin configurations $|\sigma(p)\rangle$ are readily available, as they will be required within this

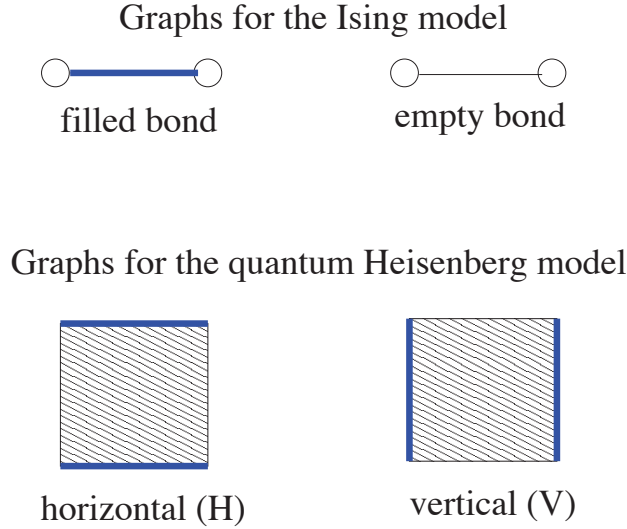


Fig. 5: Cluster components on a bond vs. graph components on a plaquette (breakups).

update scheme. In more detail, if at propagation level $p = 1, \dots, \Lambda$, an identity-operator is present, i.e., $t(p) = 0$, we attempt to replace it by a diagonal bond-operator on a bond b that is randomly chosen among all the N_b lattice bonds. We then use the Metropolis acceptance probability for this move, i.e.

$$P([0, 0] \rightarrow [1, b]) = \min \left[\frac{\beta N_b}{\Lambda - n} \langle \sigma(p-1) | H_{1,b} | \sigma(p-1) \rangle, 1 \right], \quad (42)$$

where n is the number of bond-operators in the sequence before the update. For the Heisenberg model considered here, this update can be immediately rejected, if the two spins at the chosen bond at the current propagation level are parallel, while in the other case, the matrix element simply equals $1/2$, cf. Eqs. (28)-(30). On the other hand, if the local operator at the current propagation level is a diagonal bond operator, $t(p) = 1$, an attempt is made to replace it by the identity operator with Metropolis acceptance probability

$$P([1, b] \rightarrow [0, 0]) = \min \left[\frac{\Lambda - n + 1}{\beta N_b} \langle \sigma(p-1) | H_{1,b} | \sigma(p-1) \rangle^{-1}, 1 \right]. \quad (43)$$

Once the full operator sequence has been traversed, one should have recovered the initial spin configuration, since $|\sigma(\Lambda)\rangle = |\sigma\rangle$. In case of long-ranged exchange interactions, the above update scheme can be easily generalized to ensure that the various interaction terms are efficiently sampled [28, 29].

4 The loop algorithm

We now present the improved global update schemes that have been developed during the 1990s and early 2000s for the world-line QMC approach. We will first discuss the idea behind the loop algorithm [8], based on the discrete-time formulation. An excellent, general account on the loop algorithm and related methods is provided by Evertz in his review [30].

To appreciate the idea behind the loop algorithm, it is useful to first recall the Swendsen-Wang [56] algorithm for the ferromagnetic Ising model. The goal there is to identify physically relevant clusters of parallel spins, and then collectively flip all spins within such a cluster in order to enhance the update dynamics near criticality. Essentially, one visits all bonds on the lattice and assigns a bond variable $\tau_b \in \{0, 1\}$ to each bond, by deciding if a given bond will be filled ($\tau_b = 1$) or kept empty ($\tau_b = 0$) according to certain probabilities, cf. the upper panel of Fig. 5. Namely, if the two spins connected by the bond are anti-parallel, the bond is kept empty for sure, while for parallel spins along a bond, the bond is filled with a finite probability $p = 1 - e^{-2\beta|J|}$, where $J < 0$ denotes the strength of the (ferromagnetic) Ising interaction on that bond. Then, one identifies clusters of spins connected by filled bonds, and attempts to flip each cluster individually with probability $1/2$. Within the Wolff-algorithm [57] one instead generates only one of these clusters by starting its construction from a randomly chosen lattice site, and then flips this cluster with probability 1.

For the case of the quantum model in the checkerboard decomposition, we cannot directly apply these schemes of assigning bond variables, since we need to ensure that after an update only valid, unbroken world-line configurations have been generated. Since we are given restrictions for the possible types of shaded plaquettes that can occur in valid configurations, we should thus instead of assigning bond variables assign plaquette variables. Therefore, in the loop algorithm, each plaquette P is assigned a plaquette variable (now called “graph” or “breakup”), with certain probabilities. For the antiferromagnetic Heisenberg model, there are two types of breakups that can be assigned to a plaquette; they are shown in the lower panel of Fig. 5 and we refer to them as the horizontal (H) and vertical (V) breakup. Each shaded plaquette with the state $C|_P$ is assigned the horizontal breakup (H) with probability

$$P(H) = \begin{cases} 0 & , \quad C|_P = \langle \uparrow\uparrow | e^{-\Delta\tau H_i} | \uparrow\uparrow \rangle, \langle \downarrow\downarrow | e^{-\Delta\tau H_i} | \downarrow\downarrow \rangle \\ \tanh(\frac{\Delta\tau J}{2}) & , \quad C|_P = \langle \uparrow\downarrow | e^{-\Delta\tau H_i} | \uparrow\downarrow \rangle, \langle \downarrow\uparrow | e^{-\Delta\tau H_i} | \downarrow\uparrow \rangle \\ 1 & , \quad C|_P = \langle \downarrow\uparrow | e^{-\Delta\tau H_i} | \uparrow\downarrow \rangle, \langle \uparrow\downarrow | e^{-\Delta\tau H_i} | \downarrow\uparrow \rangle \end{cases} \quad (44)$$

otherwise, it is assigned the vertical breakup (V). These graph assignment rules can be obtained upon considering an extended configuration space that combines spin and graph configurations, which we describe next. Such a general framework to describe cluster algorithms was presented by Kandel and Domany [31], generalizing the Fortuin-Kasteleyn [32] representation of the Ising model. Below we follow the formulation by Kawashima and Gubernatis [33], which makes this relation very transparent. We start from the original representation of the quantum partition function in terms of the spin (world-line) configurations,

$$Z = \sum_C W(C). \quad (45)$$

The phase space is now enlarged, by assigning a set of possible graphs G to the original configurations, such that

$$Z = \sum_C \sum_G W(C, G), \quad (46)$$

where the new weights $W(C, G) \geq 0$ are chosen as to ensure

$$\sum_G W(C, G) = W(C). \quad (47)$$

The algorithm then proceeds as follows: Given a configuration C (which implies $W(C) \neq 0$), we assign first a graph G to the configuration C , chosen with the correct probability,

$$P(G|C) = W(C, G)/W(C). \quad (48)$$

Then we choose a new configuration C' with probability $P[(C, G) \rightarrow (C', G)]$, keeping the graph G fixed. This completes a configurational update $C \rightarrow C'$ and the process repeats by choosing a new graph G' , etc. The first step, choosing graphs with probabilities $P(G|C)$, trivially obeys detailed balance. Detailed balance for the second step requires that

$$W(C, G)P[(C, G) \rightarrow (C', G)] = W(C', G)P[(C', G) \rightarrow (C, G)] \quad (49)$$

One possible solution of this detailed balance condition is provided by the heat-bath algorithm

$$P[(C, G) \rightarrow (C', G)] = \frac{W(C, G)}{W(C, G) + W(C', G)}. \quad (50)$$

The whole approach is apparently simplified a lot, if one can find an assignment of graph weights, such that $W(C, G)$ does not depend on the configuration C , whenever it is non-zero in that configuration, i.e., when $W(C, G)$ has the following form:

$$W(C, G) = \Delta(C, G)V(G), \quad (51)$$

where

$$\Delta(C, G) = \begin{cases} 1 & , \text{ if } W(C, G) \neq 0 \\ 0 & , \text{ otherwise} \end{cases}. \quad (52)$$

In this case, the heat-bath probability (50) simply becomes $P = 1/2$. Furthermore, the weight of the spin configuration $W(C)$ in fact decomposes into a product of plaquette weights, cf. Eq. (17). Further simplifications thus arise, if also the graph weight can be represented as a product of separate weights for each plaquette, i.e., if

$$W(C, G) = \prod_P W_P(C|_P, G|_P), \quad (53)$$

so that the whole graph assignment procedure can be performed locally on the level of the (shaded) plaquettes, i.e., in terms of the plaquette breakups.

While for a more general, anisotropic spin-1/2 model additional graph elements are required, the above conditions can be fulfilled for the Heisenberg antiferromagnet by employing only the two breakups (H and V) on each shaded plaquette and assigning the H breakup with the probability given in Eq. (44). An example of such an assignment of the plaquette breakups is shown in the central panel of Fig. 6. After having assigned to each plaquette a breakup, the

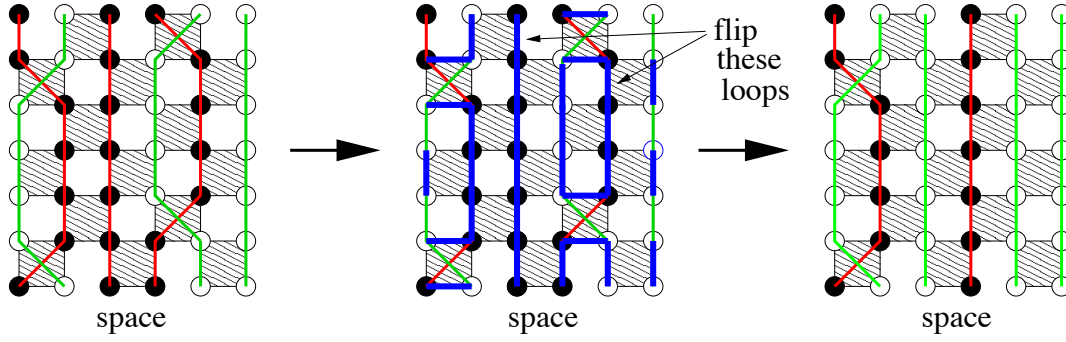


Fig. 6: An update step in the loop algorithm, where two clusters of spins are flipped.

graph edges are connected to form clusters of connected graph edges, and hence clusters (loops) of connected local spin states. Now one flips each such cluster independently with probability $P = 1/2$, thereby changing the local spins and thus the world-line configuration. An example of the whole update procedure is shown in Fig. 6. One sees from this example, that after a single loop update, one can realize large-scale changes of the world-line configuration. In particular, as seen in Fig. 6, it is possible to change the total magnetization of the system. The algorithm that we introduced above generates the whole graph, i.e., the complete set of all loops, G and flips each such cluster with probability $P = 1/2$. It is however also possible, to employ a single-loop version of the algorithm, in which, similar to the classical Wolff cluster algorithm, only a single cluster is constructed: One randomly picks a site from the world-line configuration and constructs only the cluster which includes that site. This can be done by determining the breakups (and thus the route of the loop) only on those plaquettes, that are indeed traversed during the loop construction. It is then feasible to employ the Metropolis acceptance for solving the detailed balance condition in Eq. (49), which results in the probability $P = 1$ to flip the just constructed loop. As in the classical case, the single-loop variant usually provides an even smaller dynamical critical scaling than the multi-loop variant.

Furthermore, it is also feasible to realize the loop algorithm directly in the continuous-time version of the world-line QMC formulation, in both multi- and single-loop versions [12]. Since the single-loop version paves a direct conceptual path to the worm, operator and directed loop algorithms to be introduced below, we focus here on the single-loop variant, even though it is technically a bit more involved than the multi-loop version. The key insight in realizing, that a continuous-time formulation of the loop algorithm is feasible, follows from considering the breakup probability per time, which has a continuous time limit. Indeed, for an imaginary time interval $\tau_1 < \tau_2$, during which the world-line configuration on a given bond does not change, the breakup probability is constant, and the probability density for a horizontal breakup within this imaginary time range becomes

$$\lim_{\Delta\tau \rightarrow 0} \frac{P(H)}{\Delta\tau} = \lim_{\Delta\tau \rightarrow 0} \frac{\tanh(\Delta\tau J/2)}{\Delta\tau} = \frac{J}{2}. \quad (54)$$

We are thus lead to the following procedure: To start the loop construction, a site i is picked randomly, as well as a random imaginary time τ_1 between 0 and β , from which the loop will be

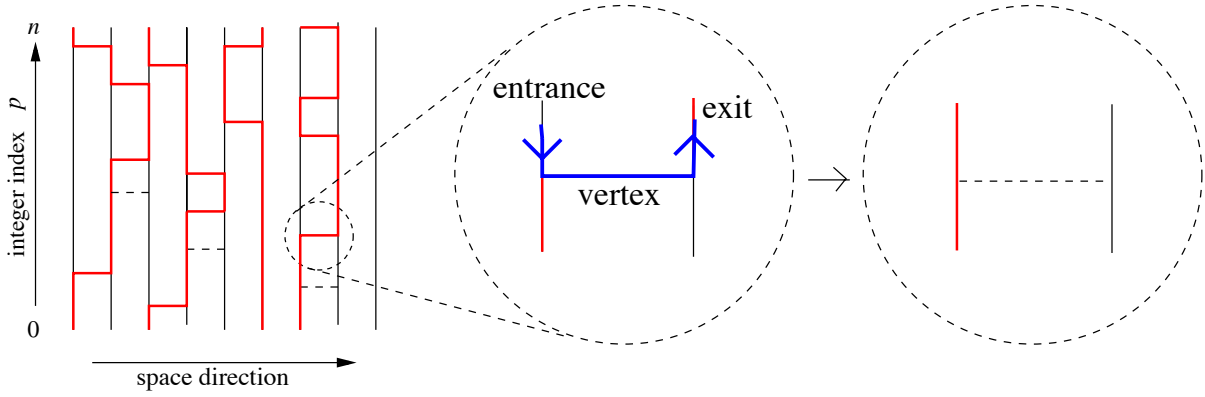


Fig. 7: A SSE configuration (left panel), the local construction of the operator loop at a 4-leg vertex corresponding to a bond operator from the SSE operator sequence (central panel), and the resulting final vertex (right panel). The path taken by the operator loop through the vertex is indicated in the central panel by the arrowed blue line (switch-and-reverse rule).

constructed, starting, e.g., initially moving upwards in time. In this forward time direction, one next identifies the time interval $\tau_1 < \tau < \tau_2$, within which the world-line configuration on all sites neighboring site i does not change. For each such neighbor j , draw a random number τ_{ij} from an exponential distribution based on the above probability density for a horizontal breakup, i.e., with $P(\tau_{ij}) \propto \exp(-J\tau_{ij}/2)$. Now, let \bar{j} be the neighbor with the smallest value of τ_{ij} , i.e., $\tau_{i\bar{j}} = \min_j(\tau_{ij})$. If $t_d = t_1 + \tau_{i\bar{j}} < t_2$, the loop-end on site i now moves up to the time t_d , and right there jumps to the site \bar{j} (a memorable analogy of this procedure is the radioactive decay with decay constant $J/2$ where the neighboring sites correspond to different decay channels). Since the constructed loops do not self-overlap, one has to exclude in the above procedure all temporal regions of neighboring sites, which have been visited by the current loop already. In case $t_d > t_2$, the loop end stays on site i and moves to time t_2 . In case the world-line jumps at time t_2 , the loop must jump as well. The whole procedure ensures the constant probability density Eq. (54) for a horizontal breakup. This process is iterated until the loop eventually closes and can be flipped as a whole.

Based on the loop algorithm, large scale simulations of up to a several million quantum spins can be performed even in (quantum) critical regions. After the original introduction of the loop algorithm for the spin-1/2 Heisenberg model in a six-vertex model formulation, it has seen various extensions to anisotropic- and higher-spin models, as well as to Hubbard and t - J models. For a detailed derivation of the breakup rules, also beyond the case of the isotropic Heisenberg model considered here, we refer to the review article by Evertz [30]. It is also possible to devise so-called “improved estimators” [34] for certain observables within the loop update, e.g., for correlation functions or magnetic susceptibilities, which provide a significant reduction of statistical errors compared to “naive” estimators for the same quantities. See Evertz’s article [30] for details.

Within the SSE formulation, which does not exhibit any systematic (Trotter) error, but still avoids the technical complications of the continuous-time formulation, it is feasible to imple-

ment a very simple version of a loop algorithm [26, 35]. This cluster update complements the local diagonal updates, that were introduced in Sec. 3.3. Namely, one considers to construct a discrete-time-like loop, based on a given SSE configuration: Consider a SSE configuration, such as the one shown in Fig. 4(c), redrawn in Fig. 7, as a collection of n vertexes (where n denotes the number of non-identity operators in the operator string), each coming along with 4 “legs”. In the central part of Fig. 7, one such 4-leg-vertex is highlighted. The loop construction starts from a randomly chosen vertex leg in the SSE configuration, called the “entrance leg”. Like in the discrete-time version of the loop algorithm, one then defines a local rule of how the loop continues through this vertex and leaves the vertex through an “exit leg”. One can interpret such rules as prescribing the scattering of the loop-head off the vertexes. The corresponding probabilities for choosing the exit leg among the four legs of the vertex need to satisfy again a local detailed balance condition. For the antiferromagnetic Heisenberg model it turns out that a very simple, deterministic scattering rule can be devised, namely, that the entrance and exit leg are always located on the same side of the vertex (the switch-and-reverse rule). This is illustrated in the central part of Fig. 7. After the exit leg has been assigned, the loop moves from this exit leg to the leg of another vertex, to which the exit leg is connected within the given SSE configuration, cf. Fig. 7. Now this leg becomes the new entrance leg on the new vertex and the whole process repeats, until the loop closes, i.e., when the last exit leg eventually equals the initial entrance leg on the starting vertex. In the example of Fig. 7, the loop closes when it reaches back to the initial entrance leg after it has visited the vertex atop the vertex from which it started. All spins along the path of the loops are then flipped. In addition to replacing all visited diagonal operators by off-diagonal operators and vice versa (cf. the right panel of Fig. 7), this operator loop update can also modify the initial spin configuration $|\sigma\rangle$, thus ensuring an ergodic sampling of the SSE configuration space if combined with the diagonal update. Of course, it is also possible in this SSE formulation to construct all the loops for a given SSE configuration and perform a multi-loop update, flipping each such loop with probability $1/2$.

5 Worms, operator loops, and directed loops

The cluster-updates that have been introduced above provide optimal performance for spin inversion symmetric Hamiltonians. However, terms in the Hamiltonian which break this symmetry, such as a magnetic field, are not taken into account during the loop construction. Instead, they would have to be included through the acceptance rate of the loop flips, which can become exponentially small at low temperatures, degrading the algorithm’s performance. Hence, it would be highly desirable to generalize the idea of a local construction of a global cluster of spins to cases, where spin-inversion symmetry is not present. The algorithms that we outline next provide such a general sampling scheme.

The worm algorithm [13] formally works in an extended configuration space, where in addition to closed world-line configurations one open world-line fragment (the “worm”) is allowed.

Formally, this is done by adding a source term to the Hamiltonian,

$$H_{\text{worm}} = H - \eta \sum_i (S_i^+ + S_i^-), \quad (55)$$

which allows world lines to be broken with a matrix element proportional to η . The worm algorithm then proceeds as follows: a worm, i.e., a world-line fragment, is created by randomly inserting a pair (S_i^+, S_i^-) of operators on a world line at nearby times in the world-line configuration. The ends of this worm are then moved randomly in space and time, using local Metropolis or heat bath updates, until the two ends of the worm meet again. Then an update which removes the worm is proposed, and if accepted we are back to a configuration with closed world lines only. This algorithm is straightforward, consisting just of local updates of the worm ends in the extended configuration space but it can perform non-local changes. Furthermore, a worm end can wind around the lattice in the temporal or spatial direction and that way change the magnetization and winding number. While not being as efficient as the loop algorithm in zero magnetic field (the worm movement follows a random walk while the loop algorithm can be interpreted as a self-avoiding random walk), the big advantage of the worm algorithm is that it remains efficient in the presence of a magnetic field. A similar algorithm was actually proposed more than a decade earlier [36]. Instead of a random walk fulfilling detailed balance at every move, the worm head in this earlier algorithm just performed a random walk. The *a posteriori* acceptance rates are then often very small and the algorithm is not efficient, just as the small acceptance rates for loop updates in magnetic fields make the loop algorithm inefficient. This highlights the importance of having the cluster-building rules of a non-local update algorithm closely tied to the physics of the problem.

Algorithms with a similar basic idea are the operator-loop update [26] in the SSE formulation and the directed-loop algorithms [27, 37] which can be formulated in both the SSE and the world-line representation. Like the worm algorithm, these algorithms create two world line discontinuities and move them around by local updates. The operator-loop algorithm for the SSE representation can be understood as a generalization of the loop algorithm for the SSE configuration, which we described at the end of Sec. 4 [26]. Again, a loop (“operator-loop”) is constructed for a given SSE configuration, starting from a random leg of a randomly chosen vertex as the entrance leg. However, for a general Hamiltonian, including, e.g., a magnetic field, the scattering rules at the vertices have to be chosen appropriately, and in general turn out to include a stochastic decision, instead of the deterministic switch-and-reverse-rule that holds for the isotropic Heisenberg model. One generic solution of the local detailed balance condition on the scattering rates is provided by a heat-bath choice among the 4 possible exit legs: the probability to choose one out of the four legs is taken to be proportional to the matrix element of the bond operator of the considered vertex in the final resulting spin configuration on this bond [26]. This choice already leads to a rather efficient algorithm. However, it is in general not excluded, that the chosen exit leg is equal to the leg on which the vertex was entered. Such a “bounce” move would result in the operator loop retracing its previous path, and thus in undoing a previously performed change to the spin configuration and the operator content of the operator string. It thus appears desirable to reduce the probability for such bounce

moves, or to even eliminate them completely. In fact, for many cases it can be shown that scattering rates can be optimized such as to eliminate completely the bounce move, while still ensuring detailed balance of the operator-loop construction [26, 27, 35, 37, 38]. An example is just the switch-and-reverse-rule for the isotropic Heisenberg antiferromagnet that was presented above. For more general models, a systematic approach to reduce the bounce probability has been formulated [27]. Such algorithms, which direct the loop away from the last changes, are called “directed loop” methods and can be formulated also for the continuous-time world-line approach [27]. Furthermore, it is possible to use linear programming techniques, in order to optimize the scattering rates with respect to the overall bounce probability for general quantum lattice models [38]. An alternative strategy to enhance the update dynamics of the operator loop algorithm is based on the concept of optimal Monte Carlo updates [39, 40]: Starting from the heat-bath solution, one iteratively improves the scattering rules by minimizing its higher eigenvalues in a matrix formulation. Both this “locally optimal solution” and the directed loops are superior to the heat-bath solution and typically perform equally well.

We furthermore want to point out that within the worm and operator-loop approaches, a natural implementation for evaluating transverse spin-spin correlation functions can be realized, based upon on-the-fly measurements of the corresponding matrix elements at the worm-ends, while constructing the worm’s path through the world-line configuration [11, 13, 24, 38]. Based on the close connection between the SSE and the continuous-time world-line approach, that we mentioned in Sec. 3.3, it is possible to implement such measurements in a rather efficient way [25]. Finally, we would like to point out that beyond the cluster update methods, which drastically reduce autocorrelation times in QMC simulations of quantum spin systems, it is also possible to adapt various extended ensemble methods such as parallel tempering [41], multicanonical methods [42, 43], and the Wang-Landau technique [44, 45], which are employed in classical Monte Carlo simulations for systems with a rough energy landscape or at first-order phase transitions, to further enhance the performance of QMC simulations in similar situations [23, 46–48].

6 The sign problem

Before closing this short review on world line QMC methods, we finally want to address a severe restriction of this approach: the sign problem. This is indeed its major limitation (note that also other unbiased QMC methods, in particular those for fermionic models, exhibit sign problems in many physically interesting situations) [49]. Currently, it seems rather unlikely that there is a general solution to this problem [50]. Let us explain what the sign problem is and where it comes from. For this purpose, consider again the weight of a plaquette in the Suzuki-Trotter decomposition that corresponds to a local spin exchange:

$$\langle \uparrow\downarrow | e^{-\Delta\tau H_i} | \downarrow\uparrow \rangle = -\sinh\left(\frac{\Delta\tau J}{2}\right). \quad (56)$$

In the antiferromagnetic case ($J > 0$), this matrix element is negative. Then we might actually worry that there will be allowed QMC world line configurations with an overall negative weight,

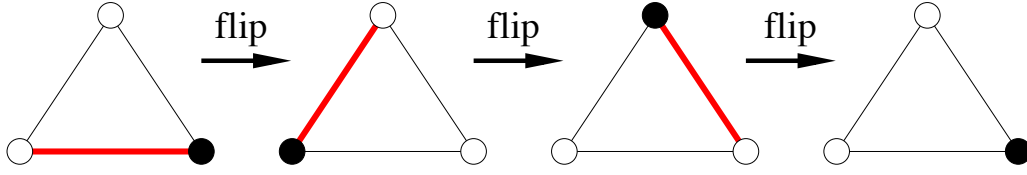


Fig. 8: Example of an odd number of spin flips on a triangle, returning to the initial state. The bond along which the spins are going to be flipped next is indicated by a red line.

which cannot be assigned as the probability for such a configuration, since probabilities ought to be positive. It is not too hard to convince yourself, that for an open chain or square lattice (in fact any bipartite lattice graph structure) the total number of spin exchange plaquettes in an allowed QMC configuration is even. In that case, the overall weight of the configuration will be positive. Hence, for the Heisenberg antiferromagnet, we are restricted to positive weights only, and can apply the Monte Carlo sampling method of the world-line configurations. However, if the lattice is not bipartite, this property is lost. To see why, consider a single triangle as an example of a non-bipartite lattice. In this case, we can start from one local spin configuration and return to the same configuration, after applying an odd number of (namely 3) spin exchanges, see Fig. 8 for an illustration. Hence, on the triangular lattice, and in fact on any frustrated, non-bipartite lattice, the weight $W(C)$ of a QMC configuration C is not necessarily positive, but can take on negative values as well. How can we then perform a MC sampling of such configurations? Well, at first sight, there seems to be an easy solution: Let us express the partition function of the quantum model in the effective classical representation as follows:

$$Z = \sum_C W(C) = \sum_C |W(C)| \cdot \text{Sgn}(C), \quad (57)$$

where we have written the weight as a product of its absolute value and its sign $\text{Sgn}(C) = W(C)/|W(C)| = \pm 1$. Now assume that we would perform a simulation of the antiferromagnet and ignore the sign, thus weighting each configuration according to the absolute values $|W(C)|$ of the weights. In fact, this precisely corresponds to simulating the ferromagnetic Heisenberg model on the same lattice, with the partition function

$$Z_F = \sum_C |W(C)|. \quad (58)$$

Using the generated configurations, we can then obtain the following expectation value of an observable A of the antiferromagnetic model,

$$\begin{aligned} \langle A \rangle &= \frac{1}{Z} \sum_C W(C) A(C) \\ &= \frac{1}{Z} \frac{Z_F}{Z_F} \sum_C |W(C)| \cdot \text{Sgn}(C) \cdot A(C) \\ &= \frac{Z_F}{Z} \frac{1}{Z_F} \sum_C |W(C)| \cdot \text{Sgn}(C) \cdot A(C), \end{aligned} \quad (59)$$

$\underbrace{\hspace{15em}}_{\langle \text{Sgn} \cdot A \rangle_F}$

from (i) the expectation value of $\langle \text{Sgn} \cdot A \rangle_F$ measured from the simulation of the ferromagnetic model (i.e., we would measure the expectation value of the product of the sign of a configuration times the value of A in that configuration), and (ii) the expectation value of the sign itself, since

$$\frac{Z}{Z_F} = \langle \text{Sgn} \rangle_F, \quad (60)$$

so that

$$\langle A \rangle = \frac{\langle \text{Sgn} \cdot A \rangle_F}{\langle \text{Sgn} \rangle_F}. \quad (61)$$

Hence, it looks like in order to obtain the expectation value of A for the antiferromagnetic model on a frustrated lattice, we could just simulate the ferromagnetic model (all weights positive), and perform the two measurements of $\langle \text{Sgn} \cdot A \rangle_F$ and $\langle \text{Sgn} \rangle_F$. In fact, such an approach works for large temperatures $T \gg J$, where the spins do not feel their exchange interactions too much, and behave weakly coupled in both the ferromagnetic and the antiferromagnetic case. But, once we are interested in the low temperature regime, then both models are characterized by very different important states. Can we really expect to learn something about the low temperature behavior of the antiferromagnet from the low temperature behavior of the ferromagnet? No, this should not be reasonable, and indeed, the above approach fails miserably at low temperatures. What happens is, that there will be almost the same number of configurations with positive weight as with negative weight, such that the average sign decreases exponentially to zero. Namely, at low temperatures we find

$$\langle \text{Sgn} \rangle_F \sim e^{-\beta N_s \Delta f}, \quad (62)$$

where Δf is the difference in the free energy per site between the antiferromagnetic and the ferromagnetic model – remember, that the free energy $F = -\ln(Z)/\beta$. Hence, in order to calculate the physical observable A from Eq. (61) we divide (in the denominator) by an exponentially small value with a finite statistical error, leading to an exponentially increasing statistical uncertainty for $\langle A \rangle$. Thus, in order to reach low temperatures and large system sizes, the computational time needs to grow exponentially and renders any useful simulation impossible. This behavior is called “the sign problem”, and makes QMC simulations of, e.g., frustrated quantum magnets practically impossible within the interesting (low-) temperature and large-system regime.

Up to date, no general solution of the QMC sign problem is known, although it can be overcome in certain special cases (cf., e.g., [35, 51, 52]). Moreover, it has been shown that a general solution to the sign problem essentially constitutes an NP-hard challenge [50]. It is however generally suspected, that no polynomial-time solutions to NP-hard problem exist. Hence, we urge the reader to contact us immediately in case she or he finds a serious path to a generic solution of the QMC sign problem! In the mean time, we hope to have stimulated your interest in performing some QMC simulations of your own for sign-problem free situations. An excellent opportunity to access application-ready codes for the methods that were introduced here is the ALPS (Algorithms and Libraries for Physics Simulations) library [53, 54]. Furthermore, basic SSE simulation codes are also available online [55]. Let the dice roll!

Appendix

A Quantum Heisenberg Model 101

For reference and to fix notation, we define the spin-1/2 quantum Heisenberg model on a finite lattice, such as a one-dimensional chain or a two-dimensional square lattice. A spin-1/2 quantum spin resides at each lattice site. Each spin \vec{S}_i , located at lattice site $i = 1, \dots, N_s$ is described by quantum mechanical spin operators

$$\vec{S}_i = \begin{pmatrix} S_i^x \\ S_i^y \\ S_i^z \end{pmatrix},$$

which fulfill the commutation relation $[S_i^\alpha, S_j^\beta] = i\hbar \varepsilon_{\alpha\beta\gamma} S_i^\gamma \delta_{ij}$, i.e., operators on different sites ($i \neq j$) commute, and locally ($i = j$) the above commutation relations reduce to the well-known algebra of spin operators. In order to set up a basis of the Hilbert space, we start from the eigenvectors of the local S_i^z operator at each lattice site i :

$$\begin{aligned} |S_i^z = +1/2\rangle &= |\uparrow\rangle_i = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_i, \\ |S_i^z = -1/2\rangle &= |\downarrow\rangle_i = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_i. \end{aligned}$$

In this local basis, the spin operators are given in matrix form as ($\hbar = 1$),

$$\begin{aligned} S_i^z &= \frac{1}{2}\sigma_z = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, & \text{i.e.} \quad \begin{aligned} S_i^z |\uparrow\rangle_i &= +\frac{1}{2} |\uparrow\rangle_i \\ S_i^z |\downarrow\rangle_i &= -\frac{1}{2} |\downarrow\rangle_i \end{aligned} \\ S_i^+ &= S_i^x + iS_i^y = \frac{1}{2}\sigma_x + i\frac{1}{2}\sigma_y = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, & \text{i.e.} \quad \begin{aligned} S_i^+ |\downarrow\rangle_i &= |\uparrow\rangle_i \\ S_i^+ |\uparrow\rangle_i &= 0 \end{aligned} \\ S_i^- &= S_i^x - iS_i^y = \frac{1}{2}\sigma_x - i\frac{1}{2}\sigma_y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, & \text{i.e.} \quad \begin{aligned} S_i^- |\uparrow\rangle_i &= |\downarrow\rangle_i \\ S_i^- |\downarrow\rangle_i &= 0 \end{aligned} \end{aligned}$$

with the spin raising and lowering operators S_i^+ and S_i^- , and the Pauli matrices

$$\begin{aligned} \sigma_x &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \sigma_z &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

For a lattice of N_s spins, we have a total of 2^{N_s} basis states of the Hilbert space,

$$\{|S_1^z, \dots, S_{N_s}^z\rangle\} = \left\{ \begin{array}{l} |\uparrow, \uparrow, \dots, \uparrow, \uparrow\rangle \\ |\uparrow, \uparrow, \dots, \uparrow, \downarrow\rangle \\ |\uparrow, \uparrow, \dots, \downarrow, \uparrow\rangle \\ \vdots \\ |\uparrow, \downarrow, \dots, \downarrow, \downarrow\rangle \\ |\downarrow, \downarrow, \dots, \downarrow, \downarrow\rangle \end{array} \right\}.$$

Each spin operator on site i acts only on the corresponding local spin. Next, we introduce an exchange coupling between nearest neighbor spins, so that the quantum spin system is described by the Hamiltonian

$$H = J \sum_{\langle i, j \rangle} \vec{S}_i \cdot \vec{S}_j,$$

where $\langle i, j \rangle$ indicates a pair of neighboring lattice sites.

In case $J > 0$, the above model is the antiferromagnetic quantum Heisenberg model, which results, e.g., within second-order perturbation theory from the half-filled Hubbard model at large local repulsions. The Hamiltonian of the Heisenberg model can also be written using spin raising and lowering operators as

$$H = J \sum_{\langle i, j \rangle} \vec{S}_i \cdot \vec{S}_j = J \sum_{\langle i, j \rangle} \left(\frac{1}{2} (S_i^+ S_j^- + S_i^- S_j^+) + S_i^z S_j^z \right).$$

Consider for example a system of only two spins (a dimer). The Hamiltonian is then given by the matrix

$$H = J \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & -\frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \text{ in the basis } \begin{array}{l} |\uparrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\uparrow\rangle \\ |\downarrow\downarrow\rangle \end{array}.$$

References

- [1] M. Suzuki, Prog. of Theor. Phys. **56**, 1454 (1976)
- [2] H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959)
- [3] M. Suzuki, S. Miyashita and A. Kuroda, Prog. Theor. Phys. **58**, 1377 (1977)
- [4] J.E. Hirsch, D.J. Scalapino, R.L. Sugar and R. Blankenbecler, Phys. Rev. Lett. **47**, 1628 (1981)
- [5] M.S. Makivić and H.-Q. Ding, Phys. Rev. B **43**, 3562 (1991)
- [6] F.F. Assaad and D. Würtz, Phys. Rev. B **44**, 2681 (1991)
- [7] N. Metropolis, A.R. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. of Chem. Phys. **21**, 1087 (1953)
- [8] H.G. Evertz, G. Lana, and M. Marcu, Phys. Rev. Lett. **70**, 875 (1993)
- [9] R.M. Fry, Phys. Rev. B **33**, 6271 (1986)
- [10] R.M. Fry and R.T. Scalettar, Phys. Rev. B **36**, 3833 (1987)
- [11] N.V. Prokof'ev, B.V. Svistunov, and I.S. Tupitsyn, Pis'ma v Zh.Eks. Teor. Fiz., **64**, 853 (1996) [English translation: cond-mat/9612091]
- [12] B.B. Beard and U.-J. Wiese, Phys. Rev. Lett. **77**, 5130 (1996)
- [13] N.V. Prokof'ev, B.V. Svistunov, and I.S. Tupitsyn, Sov. Phys. - JETP **87**, 310 (1998)
- [14] D.C. Handscomb, Proc. Cambridge Philos. Soc. **58**, 594 (1962)
- [15] D.C. Handscomb, Proc. Cambridge Philos. Soc. **60**, 115 (1964)
- [16] J.W. Lyklema, Phys. Rev. Lett. **49**, 88 (1982)
- [17] D.H. Lee, J.D. Joannopoulos, and J. W. Negele, Phys. Rev. B **30**, 1599 (1984)
- [18] E. Manousakis and R. Salvador, Phys. Rev. B **39**, 575 (1989)
- [19] A.W. Sandvik and J. Kurkijärvi, Phys. Rev. B **43**, 5950 (1991)
- [20] A.W. Sandvik, J. Phys. A **25**, 3667 (1992)
- [21] A.W. Sandvik, AIP Conf. Proc. **1297**, 135 (2010)
- [22] A. W. Sandvik, R. R. P. Singh, and D. K. Campbell, Phys. Rev. B **56**, 14510 (1997)
- [23] S. Wessel, N. Stoop, E. Gull, S. Trebst, and M. Troyer, J. Stat. Mech., P12005 (2007)

- [24] A. Dorneich and M. Troyer, Phys. Rev. E **64**, 066701 (2001)
- [25] F.F. Assaad and H.G. Evertz, in H. Fehske, R. Schneider and A. Weiße (eds.): *Computational Many-Particle Physics* (Springer, Heidelberg, 2008)
- [26] A.W. Sandvik, Phys. Rev. B **59**, R14157 (1999)
- [27] O.F. Syljuåsen and A.W. Sandvik, Phys. Rev. E **66**, 046701 (2002)
- [28] A.W. Sandvik, Phys. Rev. E **68**, 056701 (2003)
- [29] K. Fukui and S. Todo, J. Comput. Phys. **228**, 2629 (2009)
- [30] H.G. Evertz, Adv. Phys. **52**, 1 (2003)
- [31] D. Kandel and E. Domany, Phys. Rev. B **43**, 8539 (1991)
- [32] C. Fortuin and P. Kasteleyn, Physica **57**, 536 (1972)
- [33] N. Kawashima and J. Gubernatis, J. Stat. Phys. **80**, 169 (1995)
- [34] B. Ammon, H.G. Evertz, N. Kawashima, M. Troyer and B. Frischmuth, Phys. Rev. B **58**, 4304 (1998)
- [35] P. Henelius and A.W. Sandvik, Phys. Rev. B **63**, 1102 (2000)
- [36] J.J. Cullen and D.P. Landau, Phys. Rev. B **27**, 297 (1983)
- [37] O.F. Syljuåsen, Phys. Rev. E **67**, 046701 (2003)
- [38] F. Alet, S. Wessel, and M. Troyer Phys. Rev. E **71**, 036706 (2005)
- [39] L. Pollet, S.M.A. Rombouts, K. Van Houcke, and K. Heyde, Phys. Rev. E **70**, 056705 (2004)
- [40] L. Pollet, K. Van Houcke, and S.M.A. Rombouts, J. Comp. Phys. **225**, 2249 (2007)
- [41] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992)
- [42] B.A. Berg and T. Neuhaus, Phys. Lett. B. **267**, 249 (1991)
- [43] B.A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992)
- [44] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001)
- [45] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001)
- [46] P. Sengupta, A.W. Sandvik, and D.K. Campbell, Phys. Rev. B **65**, 155113 (2002)
- [47] M. Troyer, S. Wessel and F. Alet, Phys. Rev. Lett. **90**, 120201 (2003)

- [48] M. Troyer, F. Alet, and S. Wessel, *Braz. J. of Physics* **34**, 377 (2004)
- [49] E.Y. Loh Jr., J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R.L. Sugar, *Phys. Rev. B* **41**, 9301 (1990)
- [50] M. Troyer and U.-J. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005)
- [51] T. Nakamura, *Phys. Rev. B* **57**, R3197 (1998)
- [52] S. Chandrasekharan and U.-J. Wiese, *Phys. Rev. Lett.* **83**, 3116 (1999)
- [53] A. F. Albuquerque, F. Alet, P. Corboz, P. Dayal, A. Feiguin, S. Fuchs, L. Gamper, E. Gull, S. Gürtler, A. Honecker, R. Igarashi, M. Körner, A. Kozhevnikov, A. Läuchli, S. R. Manmana, M. Matsumoto, I. P. McCulloch, F. Michel, R. M. Noack, G. Pawłowski, L. Pollet, T. Pruschke, U. Schollwöck, S. Todo, S. Trebst, M. Troyer, P. Werner, and S. Wessel, *J. Magn. Mater.* **310**, 1187 (2007)
- [54] Codes available online at <http://alps.comp-phys.org>
- [55] Codes available online at [http://physics.bu.edu/\\$\sim\\$sim\\$sandvik](http://physics.bu.edu/\simsim$sandvik)
- [56] R.H. Swendsen and J.S. Wang, *Phys. Rev. Lett* **58**, 86 (1987)
- [57] U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989)

8 Quantum Theory of Molecular Magnetism

Jürgen Schnack

Bielefeld University, Faculty of Physics

Universitätsstr. 25, D-33615 Bielefeld

Contents

1	Introduction	2
2	Substances	3
3	Theoretical techniques and results	6
3.1	Hamiltonian	6
3.2	Evaluating the spectrum	7
3.3	Evaluation of thermodynamic observables	18
4	General properties of spectra	20
5	Magnetocalorics	27

1 Introduction

In this chapter I would like to present a short introduction into the field of molecular magnetism and some of the theoretical techniques as well as results. Some parts of the presentation overlap with earlier presentations in Refs. [1, 2].

The synthesis of magnetic molecules shows continuous progress for the past 20 years [3–13]. Each of the identical molecular units can contain as few as two and up to several dozens of paramagnetic ions (“spins”). One of the largest paramagnetic molecules synthesized to date, the polyoxometalate $\{\text{Mo}_{72}\text{Fe}_{30}\}$ [14] contains 30 iron ions of spin $s = 5/2$. Although these materials appear as macroscopic samples, i. e. crystals or powders, the intermolecular magnetic interactions are utterly negligible as compared to the intra-molecular interactions. Therefore, measurements of their magnetic properties reflect mainly ensemble properties of single molecules. Their magnetic features promise a variety of applications in physics, magneto-chemistry, biology, biomedicine and material sciences [3, 5, 15] as well as in quantum computing [16–19]. The most promising progress so far is being made in the field of spin crossover substances using effects like “Light Induced Excited Spin State Trapping (LIESST)” [20] as well as in the field of magnetic refrigeration [13, 21, 22]. New efforts consist in depositing magnetic molecules on surfaces and investigating their static and transport properties [23–26].

It appears that in many of these molecules – especially those which contain iron-group elements or gadolinium – the localized single-particle magnetic moments couple in most cases antiferromagnetically (and much rarer ferromagnetically), and the spectrum is rather well described by the Heisenberg model [27] with isotropic nearest neighbor interaction sometimes augmented by terms which describe various anisotropies [28, 29]. Studying such spin arrays focuses on qualitatively new physics caused by the finite (mesoscopic) size of the system.

The determination of the energy spectra of magnetic molecules is a demanding numerical problem. The attempt to diagonalize the Hamilton matrix numerically is very often severely restricted due to the huge dimension of the underlying Hilbert space. For a magnetic system of N spins of spin quantum number s the dimension is $(2s + 1)^N$ which grows exponentially with N . Group theoretical methods can help to ease this numerical problem [29–37]. With such methods one is able to block-diagonalize the Hamiltonian and thus to treat spin systems of unprecedented size. Thermodynamic observables such as the magnetization are then easily evaluated. In addition it provides a spectroscopic labeling by irreducible representations that can be related to selection rules which can be helpful when interpreting transitions induced by Electron Paramagnetic Resonance (EPR), Nuclear Magnetic Resonance (NMR) or Inelastic Neutron Scattering (INS). Besides numerically exact methods, approximations such as the Density Matrix Renormalization Group (DMRG) [38–41] and the Finite-Temperature Lanczos-Method (FTLM) [42–45] are employed for magnetic molecules.

2 Substances

From the viewpoint of theoretical magnetism it is not so important which chemical structures magnetic molecules actually have. Nevertheless, it is very interesting to note that they appear in almost all branches of chemistry. There are inorganic magnetic molecules like polyoxometalates, metal-organic molecules, and purely organic magnetic molecules in which radicals carry the magnetic moments. It is also fascinating that such molecules can be synthesized in a huge variety of structures extending from rather asymmetric structures to highly symmetric rings.

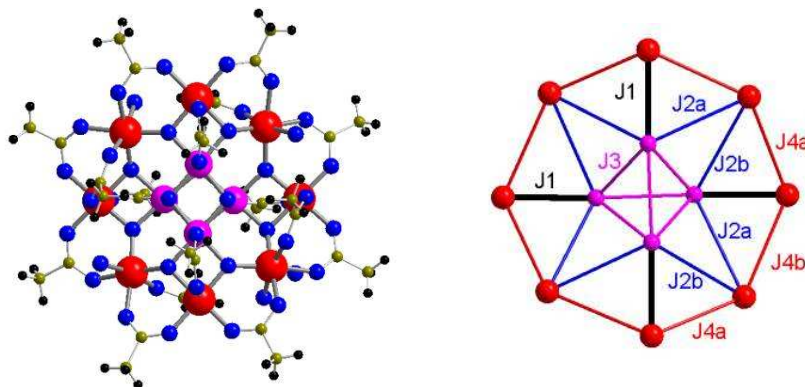


Fig. 1: Structure of Mn-12-acetate: On the l.h.s. the Mn ions are depicted by large bullets, on the r.h.s. the dominant couplings are given.

One of the first magnetic molecules to be synthesized was Mn-12-acetate [46] (Mn_{12}) – $[\text{Mn}_{12}\text{O}_{12}(\text{CH}_3\text{COO})_{16}(\text{H}_2\text{O})_4]$ – which by now serves as the “drosophila” of molecular magnetism, see e. g. [3, 15, 47–49]. As shown in Fig. 1 the molecule contains four Mn(IV) ions ($s = 3/2$) and eight Mn(III) ions ($s = 2$) which are magnetically coupled to give an $S = 10$ ground state. The molecule possesses a magnetic anisotropy, which determines the observed relaxation of the magnetization and quantum tunneling at low temperatures [47, 50].

Although the investigation of magnetic molecules in general – and of Mn-12-acetate in particular – has made great advances over the past two decades, it is still a challenge to deduce the underlying microscopic Hamiltonian, even if the Hamiltonian is of Heisenberg type. Mn-12-acetate is known for more than 20 years now and investigated like no other magnetic molecule,

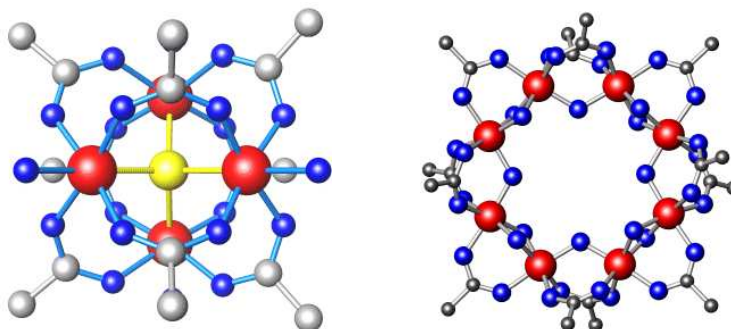


Fig. 2: Structure of a chromium-4 and a chromium-8 ring. The Cr ions are depicted in red.

but only recently its model parameters could be estimated with satisfying accuracy [51, 52]. Another very well investigated class of molecules is given by spin rings among which iron rings (“ferric wheels”) are most popular [8, 41, 53–60]. Iron-6 rings for instance can host alkali ions such as lithium or sodium which allows to modify the parameters of the spin Hamiltonian within some range [61, 62]. Another realization of rings is possible using chromium ions as paramagnetic centers [63, 64]. Figure 2 shows the structure of two rings, one with four chromium ions the other one with eight chromium ions.

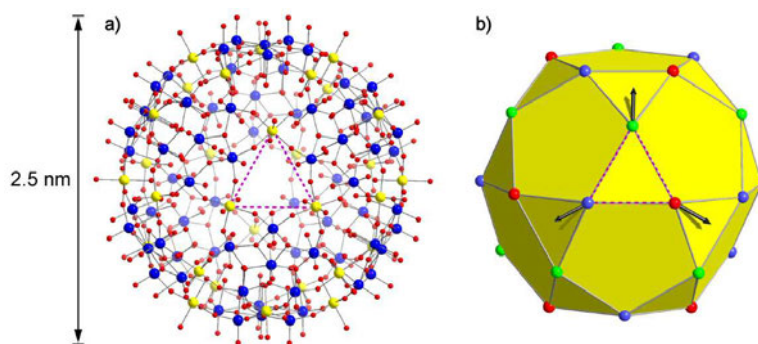


Fig. 3: Structure of $\{\text{Mo}_{72}\text{Fe}_{30}\}$, a giant Keplerate molecule where 30 iron ions are placed at the vertices of an icosidodecahedron. a) sketch of the chemical structure, b) magnetic structure showing the iron ions (spheres), the nearest neighbor interactions (edges) as well as the spin directions in the classical ground state. The dashed triangle on the left corresponds to the respective triangle on the right [66].

Another route to molecular magnetism is based on so-called Keplerate structures which allow the synthesis of truly giant and highly symmetric spin arrays [65]. The molecule $\{\text{Mo}_{72}\text{Fe}_{30}\}$ [14, 66] containing 30 iron ions of spin $s = 5/2$ may be regarded as the archetype of such structures. Figure 3 shows on the left the inner skeleton of this molecule – Fe and O-Mo-O bridges – as well as the classical ground state [67] depicted by arrows on the right [66].

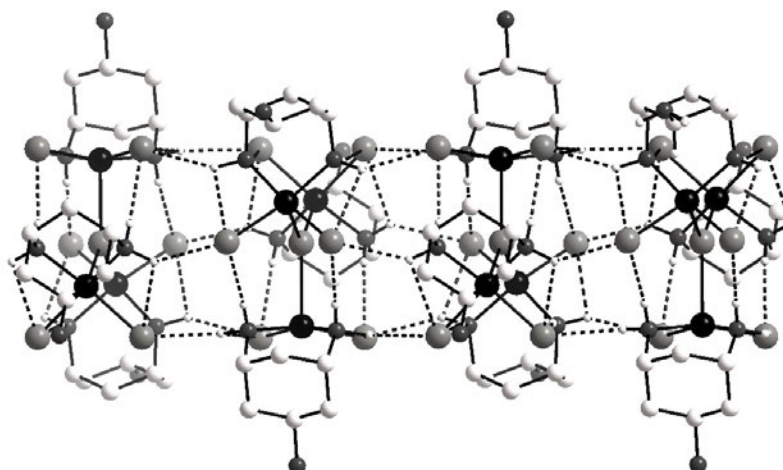


Fig. 4: Structure of a spin tube that consists of stacked Cu_3 units (black spheres) [68–70].

One of the obvious advantages of magnetic molecules is that the magnetic centers of different molecules are well separated by the ligands of the molecules. Therefore, the intermolecular interactions are utterly negligible and magnetic molecules can be considered as being independent. Nevertheless, it is also desirable to build up nanostructured materials consisting of magnetic molecules in a controlled way. Figure 4 gives an example of a linear structure consisting of stacked Cu_3 units [68,69]. These systems show new combinations of physical properties that stem from both molecular and bulk effects, the latter in the example consisting in Luttinger liquid behavior [70].

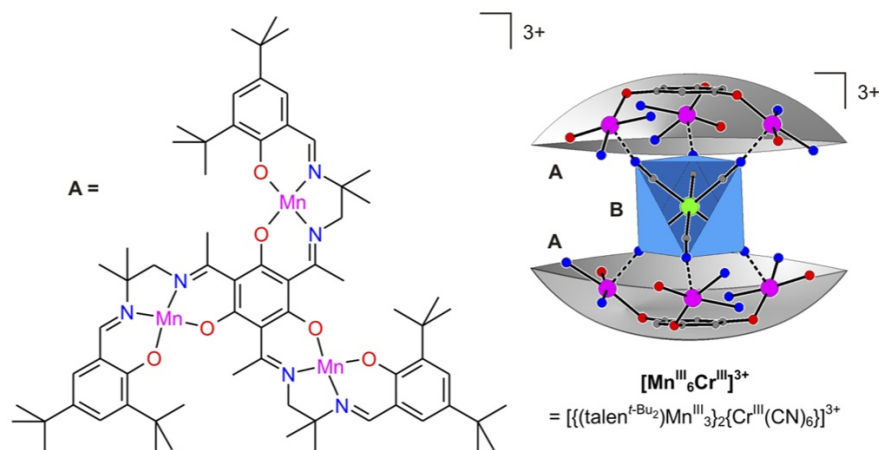


Fig. 5: Design principle of Mn_6Cr and related molecules: Two C_3 -symmetric ligands (shown on the left) enclose a central unit and thus preserve total C_3 symmetry [12, 71].

Many more structures than those sketched above can be synthesized nowadays and with the increasing success of coordination chemistry more are yet to come. The final hope of course is that magnetic structures can be designed according to the desired magnetic properties. One major step along this line is performed by the group of Thorsten Glaser in Bielefeld who synthesizes Single Molecule Magnets according to a rational design as illustrated in Fig. 5 [12,71]. Another example is given by the search for new sub-Kelvin refrigerant materials such as gadolinium containing compounds as shown in Fig. 6 [13,72].

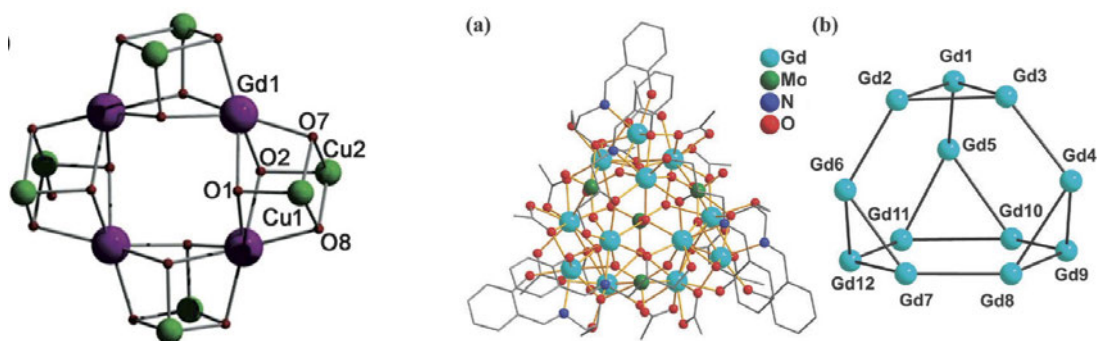


Fig. 6: Examples of Gd-containing molecules for magnetic refrigeration [13, 72].

3 Theoretical techniques and results

3.1 Hamiltonian

It appears that in the majority of these molecules the interaction between the localized single-particle magnetic moments can be rather well described by the Heisenberg model with isotropic (nearest neighbor) interaction and an additional anisotropy term [29, 30, 61, 73, 74]. Dipolar interactions are usually of minor importance. It is also found that antiferromagnetic interactions are favored in most molecules leading to nontrivial ground states.

Heisenberg Hamiltonian

For many magnetic molecules the dominant parts of the total Hamiltonian can be written as

$$\tilde{H} = \tilde{H}_{\text{Heisenberg}} + \tilde{H}_{\text{anisotropy}} + \tilde{H}_{\text{Zeeman}} \quad (1)$$

$$\tilde{H}_{\text{Heisenberg}} = - \sum_{i,j} J_{ij} \vec{s}_i \cdot \vec{s}_j \quad (2)$$

$$\tilde{H}_{\text{anisotropy}} = \sum_{i=1}^N [D_i \vec{e}_{i,3} \otimes \vec{e}_{i,3} + E_i \{ \vec{e}_{i,1} \otimes \vec{e}_{i,1} - \vec{e}_{i,2} \otimes \vec{e}_{i,2} \}] \quad (3)$$

$$\tilde{H}_{\text{Zeeman}} = \mu_B \vec{B} \cdot \sum_i \mathbf{g}_i \cdot \vec{s}_i. \quad (4)$$

The Heisenberg Hamilton operator¹ in the form given in Eq. (2) is isotropic, J_{ij} is a symmetric matrix containing the exchange parameters between spins at sites i and j . The exchange parameters are usually given in units of energy, and $J_{ij} < 0$ corresponds to antiferromagnetic, $J_{ij} > 0$ to ferromagnetic coupling.² The sum in (2) runs over all possible tuples (i, j) . The vector operators \vec{s}_i are the single-particle spin operators.

The single-ion anisotropy terms (3) usually simplify to a large extend in symmetric systems, for instance for symmetric spin rings, where the site-dependent unit vectors $\vec{e}_{i,3}$ often are all equal, and the strength as well is the same for all sites $D_i = d$ and $E_i \approx 0$.

The third part (Zeeman term) in the full Hamiltonian describes the interaction with the external magnetic field. With an isotropic and identical g -matrix the direction of the field can be assumed to be along the z -axis which simplifies the Hamiltonian very much.

Although the Hamiltonian looks rather simple, the eigenvalue problem is very often not solvable due to the huge dimension of the Hilbert space or because the number of exchange constants is too big to allow an accurate determination from experimental data. Therefore, one falls back to effective single-spin Hamiltonians for molecules with non-zero ground state spin and a large enough gap to higher-lying multiplets.

¹Operators are denoted by a tilde.

²One has to be careful with this definition since it varies from author to author

Single-spin Hamiltonian

For molecules like Mn_{12} and Fe_8 which possess a high ground state spin and well separated higher lying levels the following single-spin Hamiltonian

$$\tilde{H} = D_2 \tilde{S}_z^2 + D_4 \tilde{S}_z^4 + \tilde{H}' \quad (5)$$

$$\tilde{H}' = g\mu_B B_x \tilde{S}_x \quad (6)$$

is appropriate, see e. g. Ref. [49]. The first two terms of the Hamilton operator \tilde{H} represent the anisotropy whereas \tilde{H}' is the Zeeman term for a magnetic field along the x -axis. The total spin is fixed, i. e. $S = 10$ for Mn_{12} and Fe_8 , thus the dimension of the Hilbert space is $\dim(\mathcal{H}) = 2S + 1$.

The effective Hamiltonian (5) is sufficient to describe the low-lying spectrum and phenomena like magnetization tunneling. Since \tilde{H}' does not commute with the z -component of the total spin \tilde{S}_z , every eigenstate $|M\rangle$ of \tilde{S}_z , i. e. the states with good magnetic quantum number M , is not stationary but will tunnel through the barrier and after half the period be transformed into $|-M\rangle$.

3.2 Evaluating the spectrum

The ultimate goal is to evaluate the complete eigenvalue spectrum of the full Hamilton operator (1) as well as all eigenvectors. Since the total dimension of the Hilbert space is usually very large, e. g. $\dim(\mathcal{H}) = (2s + 1)^N$ for a system of N spins of equal spin quantum number s , a straightforward diagonalization of the full Hamilton matrix is not feasible. Nevertheless, very often the Hamilton matrix can be decomposed into a block structure because of spin symmetries or space symmetries. Accordingly the Hilbert space can be decomposed into mutually orthogonal subspaces. Then for a practical evaluation only the size of the largest matrix to be diagonalized is of importance (relevant dimension).

Product basis

The starting point for any diagonalization is the product basis $|\vec{m}\rangle = |m_1, \dots, m_j, \dots, m_N\rangle$ of the single-particle eigenstates of all \tilde{S}_j^z

$$\tilde{S}_j^z |m_1, \dots, m_j, \dots, m_N\rangle = m_j |m_1, \dots, m_j, \dots, m_N\rangle. \quad (7)$$

These states are sometimes called Ising states. They span the full Hilbert space and are used to construct symmetry-related basis states.

Symmetries of the problem

Since the isotropic Heisenberg Hamiltonian includes only a scalar product between spins, this operator is rotationally invariant in spin space, i. e. it commutes with $\vec{\tilde{S}}$ and thus also with

$$[\tilde{H}_{\text{Heisenberg}}, \tilde{S}^2] = 0 \quad , \quad [\tilde{H}_{\text{Heisenberg}}, \tilde{S}_z] = 0. \quad (8)$$

In a case where anisotropy is negligible a well-adapted basis is thus given by the simultaneous eigenstates $|S, M, \alpha\rangle$ of \vec{S}^2 and S_z , where α enumerates those states belonging to the same S and M [33,36,37,75]. Since the applied magnetic field can be assumed to point into z -direction for vanishing anisotropy the Zeeman term automatically also commutes with $\underline{H}_{\text{Heisenberg}}$, \vec{S}^2 , and S_z . Since M is a good quantum number the Zeeman term does not need to be included in the diagonalization but can be added later.

Besides spin symmetries many molecules possess spatial symmetries. One example is given by spin rings which have a translational symmetry. In general the symmetries depend on the point group of the molecule; for the evaluation of the eigenvalue spectrum its irreducible representations have to be used [30,33,36,37,61]. Thus, in a case with anisotropy one loses spin rotational symmetries but one can still use space symmetries. Without anisotropy one even gains a further reduction of the relevant dimension.

Dimension of the problem

The following section illuminates the relevant dimensions assuming certain symmetries [75]. If no symmetry is present the total dimension is just

$$\dim(\mathcal{H}) = \prod_{j=1}^N (2s_j + 1) \quad (9)$$

for a spin array of N spins with various spin quantum numbers. In many cases the spin quantum numbers are equal resulting in a dimension of the total Hilbert space of $\dim(\mathcal{H}) = (2s + 1)^N$. If the Hamiltonian commutes with S_z then M is a good quantum number and the Hilbert space \mathcal{H} can be divided into mutually orthogonal subspaces $\mathcal{H}(M)$

$$\mathcal{H} = \bigoplus_{M=-S_{\max}}^{+S_{\max}} \mathcal{H}(M), \quad S_{\max} = \sum_{j=1}^N s_j. \quad (10)$$

For given values of M , N and of all s_j the dimension $\dim(\mathcal{H}(M))$ can be determined as the number of product states (7), which constitute a basis in $\mathcal{H}(M)$, with $\sum_j m_j = M$. The solution of this combinatorial problem can be given in closed form [75]

$$\dim(\mathcal{H}(M)) = \frac{1}{(S_{\max} - M)!} \left[\left(\frac{d}{dz} \right)^{S_{\max} - M} \prod_{j=1}^N \frac{1 - z^{2s_j+1}}{1 - z} \right]_{z=0}. \quad (11)$$

For equal single-spin quantum numbers s , and thus a maximum total spin quantum number of $S_{\max} = Ns$, (11) simplifies to

$$\begin{aligned} \dim(\mathcal{H}(M)) &= f(N, 2s + 1, S_{\max} - M) \quad \text{with} \\ f(N, \mu, \nu) &= \sum_{n=0}^{\lfloor \nu/\mu \rfloor} (-1)^n \binom{N}{n} \binom{N - 1 + \nu - n\mu}{N - 1}. \end{aligned} \quad (12)$$

In both formulae (11) and (12), M may be replaced by $|M|$ since the dimension of $\mathcal{H}(M)$ equals those of $\mathcal{H}(-M)$. $\lfloor \nu/\mu \rfloor$ in the sum symbolizes the greatest integer less or equal to ν/μ . Eq. (12) is known as a result of de Moivre [76].

If the Hamiltonian commutes with \vec{S}^2 and all individual spins are identical the dimensions of the orthogonal eigenspaces $\mathcal{H}(S, M)$ can also be determined. The simultaneous eigenspaces $\mathcal{H}(S, M)$ of \vec{S}^2 and S_z are spanned by eigenvectors of \vec{H} . The one-dimensional subspace $\mathcal{H}(M = S_{\max}) = \mathcal{H}(S_{\max}, S_{\max})$, especially, is spanned by $|\Omega\rangle$, a state called magnon vacuum. The total ladder operators (spin rising and lowering operators) are

$$S^\pm = S_x \pm i S_y . \quad (13)$$

For $S > M$, S^- maps any normalized \vec{H} -eigenstate $\in \mathcal{H}(S, M + 1)$ onto an \vec{H} -eigenstate $\in \mathcal{H}(S, M)$ with norm $\sqrt{S(S + 1) - M(M + 1)}$.

For $0 \leq M < S_{\max}$, $\mathcal{H}(M)$ can be decomposed into orthogonal subspaces

$$\mathcal{H}(M) = \mathcal{H}(M, M) \oplus S^- \mathcal{H}(M + 1) \quad (14)$$

with

$$S^- \mathcal{H}(M + 1) = \bigoplus_{S \geq M+1} \mathcal{H}(S, M) . \quad (15)$$

In consequence, the diagonalization of \vec{H} in \mathcal{H} has now been traced back to diagonalization in the subspaces $\mathcal{H}(S, S)$, the dimension of which are for $S < S_{\max}$

$$\dim(\mathcal{H}(S, S)) = \dim(\mathcal{H}(M = S)) - \dim(\mathcal{H}(M = S + 1)) \quad (16)$$

and can be calculated according to (12).

The use of combined spin, i.e. $SU(2)$, and point group symmetries is rather involved, compare [36, 37, 77]. I would therefore like to demonstrate the use of symmetries for the combined use of the S_z -symmetry and the translational symmetry which is a cyclic group.

As a special example I focus on the translational symmetry found in spin rings. The discussed formalism can as well be applied to other cyclic symmetries. Any such translation is represented by the cyclic shift operator \vec{T} or a multiple repetition. \vec{T} is defined by its action on the product basis (7)

$$\vec{T} |m_1, \dots, m_{N-1}, m_N\rangle = |m_N, m_1, \dots, m_{N-1}\rangle . \quad (17)$$

The eigenvalues of \vec{T} are the N -th roots of unity

$$z_k = \exp \left\{ -i \frac{2\pi k}{N} \right\} , \quad k = 0, \dots, N - 1 , \quad p_k = 2\pi k / N , \quad (18)$$

where k will be called translational (or shift) quantum number and p_k momentum quantum number or crystal momentum. The numbers z_k are called characters of the point group C_N .

The shift operator \tilde{T} commutes not only with the Hamiltonian but also with total spin. Any $\mathcal{H}(S, M)$ can therefore be decomposed into simultaneous eigenspaces $\mathcal{H}(S, M, k)$ of \tilde{S}^2 , \tilde{S}_z and \tilde{T} .

In the following we demonstrate how an eigenbasis of both \tilde{S}_z and \tilde{T} can be constructed, this basis spans the orthogonal Hilbert spaces $\mathcal{H}(M, k)$. A special decomposition of \mathcal{H} into orthogonal subspaces can be achieved by starting with the product basis and considering the equivalence relation

$$|\psi\rangle \cong |\phi\rangle \Leftrightarrow |\psi\rangle = \tilde{T}^n |\phi\rangle, \quad n \in \{1, 2, \dots, N\} \quad (19)$$

for any pair of states belonging to the product basis. The equivalence relation then induces a complete decomposition of the basis into disjoint subsets, i.e. the equivalence classes. A “cycle” is defined as the linear span of such an equivalence class of basis vectors. The obviously orthogonal decomposition of \mathcal{H} into cycles is compatible with the decomposition of \mathcal{H} into the various $\mathcal{H}(M)$. Evidently, the dimension of a cycle can never exceed N . Cycles are called “proper cycles” if their dimension equals N , they are termed “epicycles” else. One of the N primary basis states of a proper cycle may arbitrarily be denoted as

$$|\psi_1\rangle = |m_1, \dots, m_N\rangle \quad (20)$$

and the remaining ones may be enumerated as

$$|\psi_{n+1}\rangle = \tilde{T}^n |\psi_1\rangle, \quad n = 1, 2, \dots, N-1. \quad (21)$$

The cycle under consideration is likewise spanned by the states

$$|\chi_k\rangle = \frac{1}{\sqrt{N}} \sum_{\nu=0}^{N-1} \left(e^{i\frac{2\pi k}{N}} \tilde{T} \right)^\nu |\psi_1\rangle \quad (22)$$

which are eigenstates of \tilde{T} with the respective shift quantum number k . Consequently, every k occurs once in a proper cycle. An epicycle of dimension D is spanned by D eigenstates of \tilde{T} with each of the translational quantum numbers $k = 0, N/D, \dots, (D-1)N/D$ occurring exactly once. As a rule of thumb one can say that the dimension of each $\mathcal{H}(M, k)$ is approximately $\dim(\mathcal{H}(M, k)) \approx \dim(\mathcal{H}(M))/N$. An exact evaluation of the relevant dimensions for spin rings can be obtained from Ref. [75].

Exact diagonalization

If the relevant dimension is small enough, the respective Hamilton matrices can be diagonalized, either analytically [75, 78, 79] or numerically, see e.g. [30, 33, 80–85].

Again, how such a project is carried out, will be explained with the help of an example, a simple spin ring with $N = 6$ and $s = 1/2$. The total dimension is $\dim(\mathcal{H}) = (2s + 1)^N = 64$. The Hamilton operator (2) simplifies to

$$\tilde{H}_{\text{Heisenberg}} = -2J \sum_{j=1}^N \vec{s}_j \cdot \vec{s}_{j+1}, \quad N+1 \equiv 1. \quad (23)$$

We start with the magnon vacuum $|\Omega\rangle = |+++++\rangle$ which spans the Hilbert space $\mathcal{H}(M)$ with $M = N_s = 3$. “ \pm ” are shorthand notations for $m = \pm 1/2$. The dimension of the subspace $\dim(\mathcal{H}(M = N_s))$ is one and the energy eigenvalue is $E_\Omega = -2JN_s^2 = -3J$. $|\Omega\rangle$ is an eigenstate of the shift operator with $k = 0$. Since S is also a good quantum number in this example $|\Omega\rangle$ has to be an eigenstate of \tilde{S}^2 , too, the quantum number is $S = N_s$.

The next subspace $\mathcal{H}(M)$ with $M = N_s - 1 = 2$ is spanned by $|-++++\rangle$ and the five other vectors which are obtained by repetitive application of \tilde{T} . This subspace obviously has the dimension N , and the cycle spanned by $\tilde{T}^n |-++++\rangle, n = 0, \dots, N - 1$ is a proper one. Therefore, each k quantum number arises once. The respective eigenstates of \tilde{T} can be constructed according to Eq. (22) as

$$|M = 2, k\rangle = \frac{1}{\sqrt{N}} \sum_{\nu=0}^{N-1} \left(e^{i\frac{2\pi k}{N}} \tilde{T} \right)^\nu |-++++\rangle. \quad (24)$$

All subspaces $\mathcal{H}(M, k)$ have dimension one. Since $\tilde{S}^- |\Omega\rangle$ is a state belonging to $\mathcal{H}(M = N_s - 1)$ with the same k -quantum number as $|\Omega\rangle$ it is clear that $|M = 2, k = 0\rangle$ is already an eigenstate of \tilde{S}^2 with $S = N_s$. The other five $|M = 2, k \neq 0\rangle$ must have $S = N_s - 1$.

The next subspace $\mathcal{H}(M)$ with $M = N_s - 2 = 1$ is spanned by three basic vectors, i.e. $|- - + + + \rangle, |- + - + + \rangle, |- + + - + \rangle$ and the repetitive applications of \tilde{T} onto them. The first two result in proper cycles, the third vector $|- + + - + \rangle$ results in an epicycle of dimension three, thus for the epicycle we find only k quantum numbers $k = 0, 2, 4$. The energy eigenvalues found in the subspace $\mathcal{H}(M = N_s - 1)$ (“above”) must reappear here which again allows to address an S quantum number to these eigenvalues. The dimension of the subspace $\mathcal{H}(M = 1)$ is 15, the dimensions of the subspaces $\mathcal{H}(M, k)$ are 3 ($k = 0$), 2 ($k = 1$), 3 ($k = 2$), 2 ($k = 3$), 3 ($k = 4$), and 2 ($k = 5$).

The last subspace to be considered belongs to $M = 0$ and is spanned by $|- - - + + \rangle, |- - + - + \rangle, |- + - - + \rangle, |- + - + - \rangle$ and repetitive applications of \tilde{T} . Its dimension is 20. Here $|- + - + - \rangle$ leads to an epicycle of dimension two.

The Hamilton matrices in subspaces with $M < 0$ need not be diagonalized due to the \tilde{S}_z -symmetry, i.e. eigenstates with negative M can be obtained by transforming all individual $m_j \rightarrow -m_j$. Summing up the dimensions of all $\mathcal{H}(M)$ yields $1+6+15+20+15+6+1 = 64 \sqrt{}$. Figure 7 shows the resulting energy spectrum both as a function of total spin S and as a function of translational quantum number k .

Lanczos method

Complex hermitian matrices can be completely diagonalized numerically on a normal PC up to a size of about 10,000 by 10,000 which corresponds to about 1.5 Gigabyte of necessary RAM. Nevertheless, for larger systems one can still use numerical methods to evaluate low-lying energy levels and the respective eigenstates with high accuracy. Very accurate methods rest on the construction of a so-called Krylov space. One of these methods to partially diagonalize a huge matrix was proposed by Cornelius Lanczos in 1950 [86,87]. This method uses a (random)

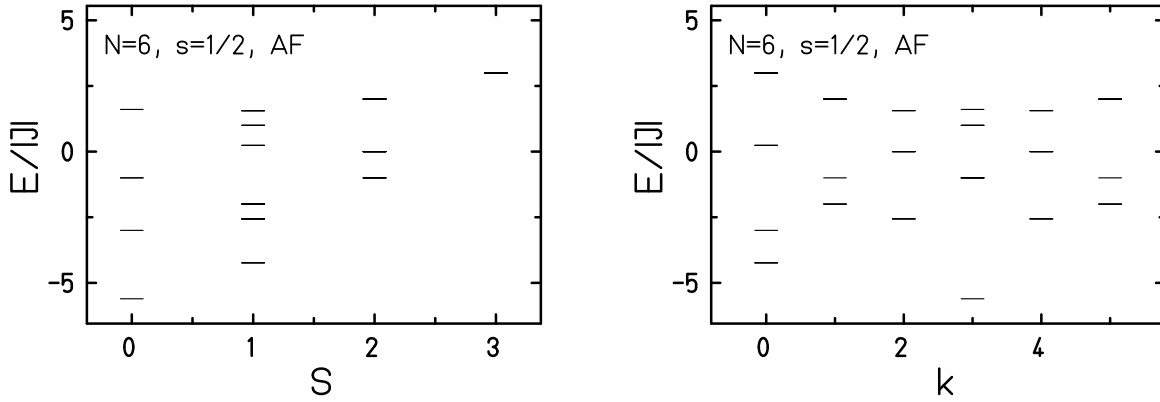


Fig. 7: Energy eigenvalues as a function of total spin quantum number S (l.h.s.) and k (r.h.s.).

initial vector. It then generates an orthonormal system in such a way that the representation of the operator of interest is tridiagonal. Every iteration adds one row and one column to the tridiagonal matrix. With growing size of the matrix its eigenvalues converge against the true ones until, in the case of finite dimensional Hilbert spaces, the eigenvalues reach their true values. The key point is that the extremal eigenvalues converge rather quickly compared to the other ones [88]. Thus, after about 100 Lanczos steps the ground state energy is already approximated to 10 figures although the dimension of the underlying Hilbert space is 10^8 .

A simple Lanczos algorithm looks like the following. One starts with an arbitrary vector $|\psi_0\rangle$, which has to have an overlap with the (unknown) ground state. The next orthogonal vector is constructed by application of \tilde{H} and projecting out the original vector $|\psi_0\rangle$

$$|\psi'_1\rangle = (1 - |\psi_0\rangle\langle\psi_0|) \tilde{H} |\psi_0\rangle = \tilde{H} |\psi_0\rangle - \langle\psi_0|\tilde{H}|\psi_0\rangle |\psi_0\rangle, \quad (25)$$

which yields the normalized vector

$$|\psi_1\rangle = \frac{|\psi'_1\rangle}{\sqrt{\langle\psi'_1|\psi'_1\rangle}}. \quad (26)$$

Similarly all further basis vectors are generated

$$\begin{aligned} |\psi'_{k+1}\rangle &= (1 - |\psi_k\rangle\langle\psi_k| - |\psi_{k-1}\rangle\langle\psi_{k-1}|) \tilde{H} |\psi_k\rangle \\ &= \tilde{H} |\psi_k\rangle - \langle\psi_k|\tilde{H}|\psi_k\rangle |\psi_k\rangle - \langle\psi_{k-1}|\tilde{H}|\psi_k\rangle |\psi_{k-1}\rangle \end{aligned} \quad (27)$$

and

$$|\psi_{k+1}\rangle = \frac{|\psi'_{k+1}\rangle}{\sqrt{\langle\psi'_{k+1}|\psi'_{k+1}\rangle}}. \quad (28)$$

The new Lanczos vector is by construction orthogonal to the two previous ones. Without proof we repeat that it is then also orthogonal to all other previous Lanczos vectors. This constitutes the tridiagonal form of the resulting Hamilton matrix

$$T_{i,j} = \langle\psi_i|\tilde{H}|\psi_j\rangle \quad \text{with} \quad T_{i,j} = 0 \quad \text{if} \quad |i-j| > 1. \quad (29)$$

The Lanczos matrix T can be diagonalized at any step. Usually one iterates the method until a certain convergence criterion is fulfilled.

The eigenvectors of \tilde{H} can be approximated using the eigenvectors ϕ_μ of T

$$|\chi_\mu\rangle \approx \sum_{i=0}^n \phi_{\mu,i} |\psi_i\rangle, \quad (30)$$

where μ labels the desired energy eigenvalue, e.g. the ground state energy. n denotes the number of iterations, and $\phi_{\mu,i}$ is the i -th component of the n -dimensional eigenvector ϕ_μ of T .

The simple Lanczos algorithm has some problems due to limited accuracy. One problem is that eigenvalues may collapse. Such problems can be solved with more refined formulations of the method [87].

Density Matrix Renormalization Group Method

The DMRG technique [38] has become one of the standard numerical methods for quantum lattice calculations in recent years [40]. Its basic idea is the reduction of Hilbert space while focusing on the accuracy of a target state. For this purpose the system is divided into subunits – blocks – which are represented by reduced sets of basis states. The dimension m of the truncated block Hilbert space is a major input parameter of the method and to a large extent determines its accuracy. For further details please see the lecture of Ulrich Schollwöck.

DMRG is best suited for chain-like structures. Many accurate results have been achieved by applying DMRG to various (quasi-)one-dimensional systems [84, 89, 90]. The best results were found for the limit of infinite chains with open boundary conditions. It is commonly accepted that DMRG reaches maximum accuracy when it is applied to systems with a small number of interactions between the blocks, e.g. systems with only nearest-neighbor interaction [91].

It is not *a priori* clear how good results for finite systems like magnetic molecules are. Such systems are usually not chain-like, so in order to carry out DMRG calculations a mapping onto a one-dimensional structure has to be performed [91]. Since the spin array consists of a countable number of spins, any arbitrary numbering is already a mapping onto a one-dimensional structure. However, even if the original system had only nearest-neighbor exchange, the new one-dimensional system has many long-range interactions depending on the way the spins are enumerated. Therefore, a numbering which minimizes long range interactions is preferable.

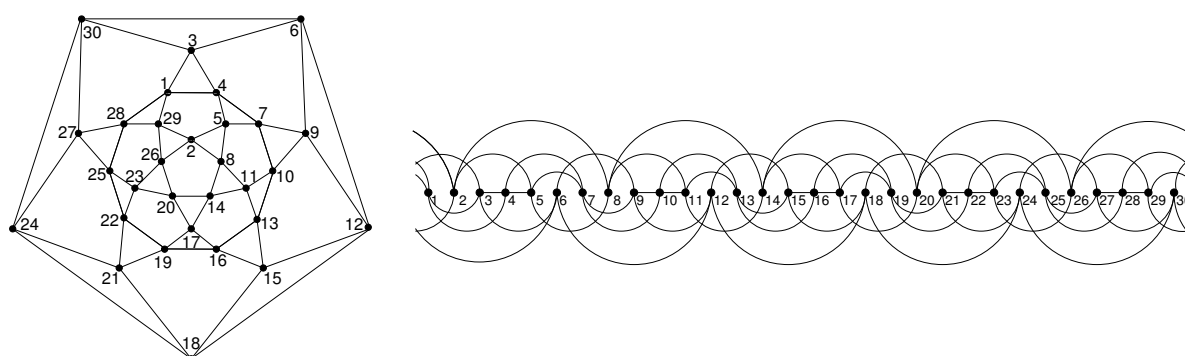


Fig. 8: Two-dimensional and one-dimensional projection of the icosidodecahedron, the site numbers are those used in [39], the lines represent interactions.

Fig. 8 shows the graph of interactions for the molecule $\{\text{Mo}_{72}\text{Fe}_{30}\}$ which was used in Ref. [39]. For finite systems a block algorithm including sweeps, which is similar to the setup in White’s original article [38], has turned out to be most efficient.

For illustrative purposes we use a simple Heisenberg Hamiltonian, compare (2). The Hamiltonian is invariant under rotations in spin space. Therefore, the total magnetic quantum number M is a good quantum number and we can perform our calculation in each orthogonal subspace $\mathcal{H}(M)$ separately [41,92].

When DMRG is applied to spin systems that are not one-dimensional, the usual way is to map the system on a one-dimensional chain with long-range interactions, i.e., to number the spins of the lattice [93]. However, if not very simple systems such as, e.g., ladders are investigated, it is not clear, which numbering is best suited. Such a problem also occurs, when DMRG is applied in the context of quantum chemistry, where models similar to the Hubbard model with long-range interactions appear and the ordering, i.e., the numbering of the orbitals is relevant. Since long-range interactions diminish the accuracy of DMRG (cf. Ref. [94]) it is clear that a good ordering needs to minimize such long-range interactions.

We have tested several numberings for the icosidodecahedron. The resulting coupling matrices J_{ij} are shown in Fig. 9. The numbering used by Exler and Schnack in an earlier investigation [39] (see top left of Fig. 9) gives a very regular “interaction pattern” with rather-short-ranged interactions, but the “periodic boundaries”, i.e., interactions between the first and the last spins, are clearly not optimal for the DMRG algorithm with two center sites. As proposed in Ref. [95],

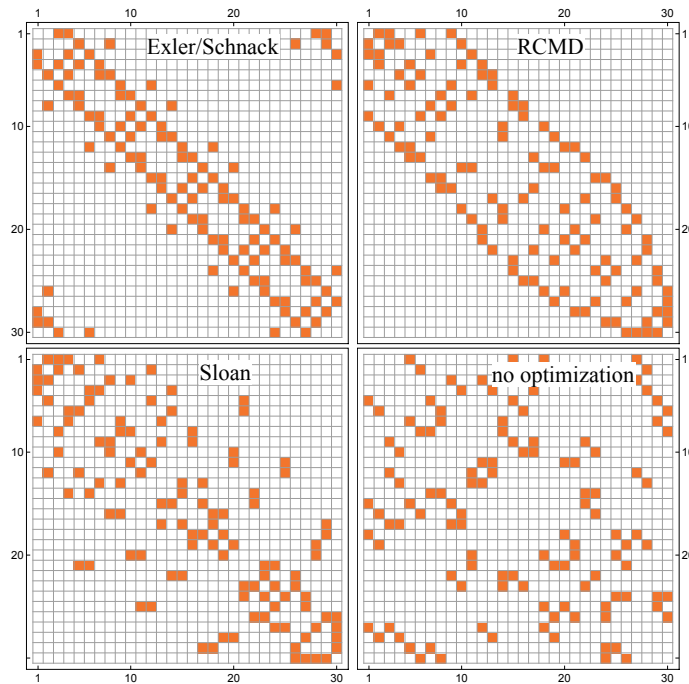


Fig. 9: Depiction of the coupling matrix (J_{ij}) for four different numberings of the vertices of the icosidodecahedron. Nonzero entries are denoted by the orange squares. Top left: numbering according to Exler and Schnack (see Ref. [39]); top right: result of the RCMD algorithm; bottom left: result of the Sloan algorithm; bottom right: unoptimized numbering.

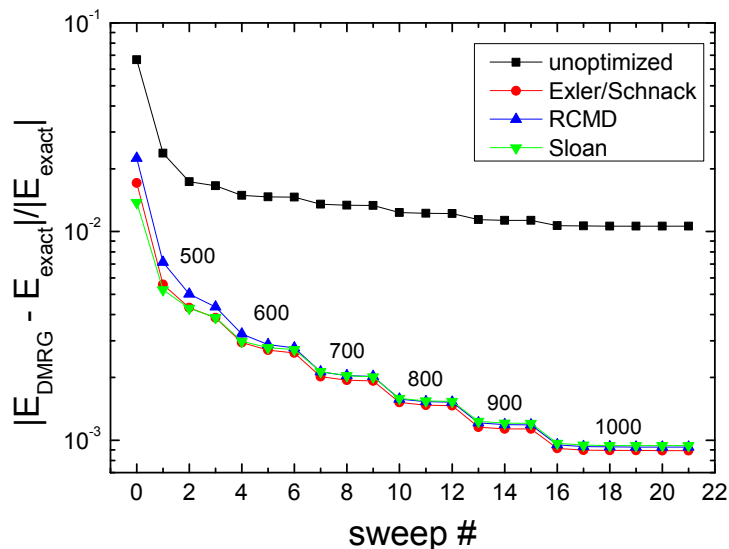


Fig. 10: DMRG results for different numberings (see Fig. 9) of the spins sitting on the vertices of the icosidodecahedron. The plot shows the error in the ground state energy as a function of the number of DMRG sweeps. The numbers above the symbols denote the number of kept density matrix eigenstates for the sweep [92].

we have used a variant of the reverse Cuthill-McKee algorithm [96, 97], the RCMD algorithm, which aims to number the vertices such that the bandwidth of the matrix is minimized. We have also used the Sloan algorithm [98] which minimizes the “envelope size”, i.e., the sum of the “row bandwidths”. (The bandwidth is the maximum of the row bandwidths.) The figure also shows an unoptimized numbering for comparison.

The results of DMRG calculations (using the ALPS DMRG code [99]) for the different spin numberings are shown in Fig. 10. We have calculated the ground state energy of the $s = 1/2$ icosidodecahedron with an increasing number of kept density matrix eigenstates (m) so that the convergence can be investigated and a comparison with the exact ground state energy (see Ref. [100]) is possible. One can see that the different optimized numberings (Exler/Schnack, RCMD, and Sloan) give almost identical results whereas the unoptimized numbering gives much worse results. These results show that a “good” numbering of the spins is absolutely essential if the DMRG method is applied to a spin system with a complicated structure. For the following results we have always used the numbering as proposed by Exler and Schnack.

As a next step we have calculated the lowest energies in the M subspaces for the icosidodecahedron with $s > 1/2$ using DMRG. The results for the $s = 1/2$ system already showed that DMRG is able to produce very accurate results for this system with relative errors smaller than 10^{-3} [39, 92].

Figure 11 shows the lowest energy eigenvalues in the subspaces of total magnetic quantum number M for the icosidodecahedron with $s = 1$ and $s = 3/2$ as obtained by DMRG and – for the large- M subspaces ($M > 18$ for $s = 1$ and $M > 33$ for $s = 3/2$) – Lanczos calculations. We have used up to $m = 2500$ density matrix eigenstates for the $s = 1$ case and up to $m = 2000$ for the $s = 3/2$ case. The largest truncated weight within a sweep is of the order of $7 \cdot 10^{-4}$ for the $M = 0$ subspace of the $s = 1$ icosidodecahedron and of the order of $4 \cdot 10^{-4}$ for the

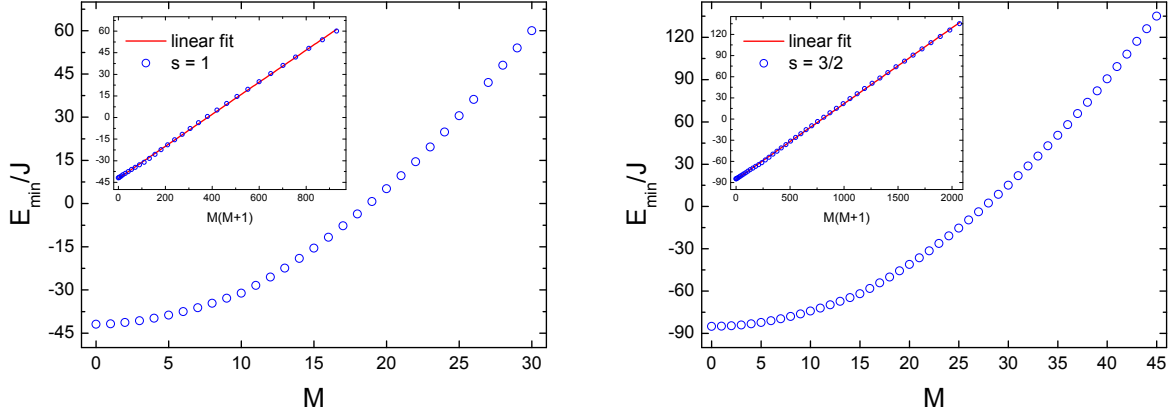


Fig. 11: Lowest energy eigenvalues in the subspaces of total magnetic quantum number M as obtained by DMRG calculations [92]. The eigenvalues for the smallest subspaces, i.e., for large M , were calculated using the Lanczos algorithm. For the DMRG calculations, the ALPS DMRG code was employed [99]. For the $s = 1$ system we have kept $m = 2500$ density matrix eigenstates in all DMRG calculations. For the $s = 3/2$ system we have kept $m = 2000$ states for the subspaces up to $M = 4$, $m = 1500$ states for the subspaces $5 \leq M \leq 23$, and 1000 states for the subspaces $M > 23$.

$s = 3/2$ case. That the truncated weight for the $s = 1$ icosidodecahedron is larger than for $s = 3/2$ although more states have been used for $s = 1$ indicates that it cannot be reliably used for a quantitative estimate of the error. The reason for this behavior might be that the results are not yet fully converged for the value of m that we have used, although we have carried out up to 60 sweeps for the calculations.

The rotational band model, compare next section, predicts a behavior of the form $E_{\min}(M) = aM(M + 1) + b$, i.e., a parabolic dependence [101]. The insets of Fig. 11 show that this is an overall good approximation for the energy eigenvalues of the full Heisenberg model. The simple rotational band approximation predicts a proportionality constant of $a = 0.1$. The linear fits as shown in the insets give the results $a = 0.111$ for $s = 1$ and $a = 0.108$ for $s = 3/2$, very close to the simple rotational band approximation. However, if one uses these (DMRG) data to

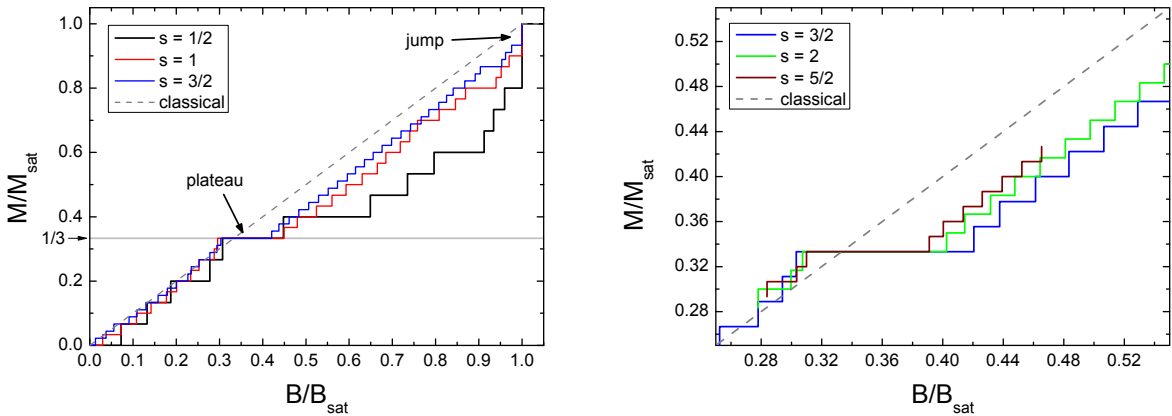


Fig. 12: Zero-temperature magnetization curves of the icosidodecahedron as obtained by DMRG calculations (cf. Fig. 11). The data for $s = 1/2$ is taken from Ref. [100], the classical result from [67]. The data is normalized to the saturation field and magnetization. L.h.s.: $s = 1/2, 1, 3/2$ and classical result. R.h.s.: $s = 3/2, 2, 5/2$ and classical result [92].

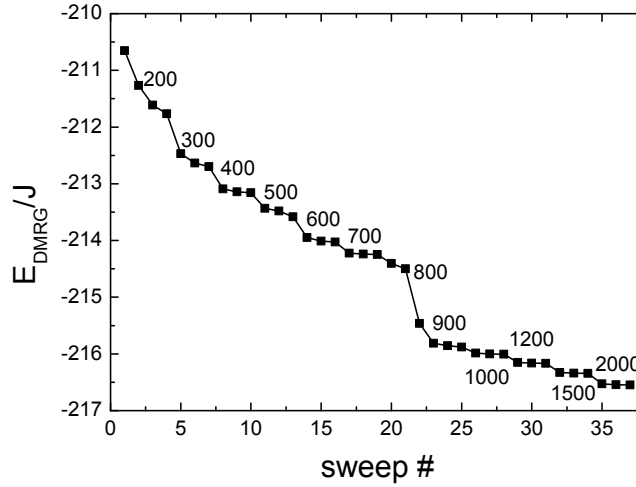


Fig. 13: Ground state energy of the $s = 5/2$ icosidodecahedron as a function of the DMRG sweep. The numbers show the retained density matrix eigenstates for the current sweep [92].

calculate the zero-temperature magnetization curve, it becomes clear that there are some crucial deviations from the ideal parabolic dependence. If there was an ideal parabolic dependence, the resulting magnetization curve would consist of steps with constant widths. Fig. 12 shows the resulting zero-temperature magnetization curves as calculated using the DMRG data.

One can see, that the magnetization curves do not consist of steps with constant widths. There are some anomalies as expected for frustrated systems. The plateaus at $\mathcal{M}/\mathcal{M}_{\text{sat}} = 1/3$ are clearly visible. The magnetization jumps due to the independent magnons [102] are also visible. Since the jump has a height of $\Delta M = 3$, it is clear that this effect vanishes for $s \rightarrow \infty$ in the plot because the magnetization is normalized by the saturation magnetization $\mathcal{M}_{\text{sat}} = 30g\mu_B s$. For $s \rightarrow \infty$, the classical result, i.e., a strictly linear magnetization curve [67], will be reached. On the contrary, the plateau seems to be very stable when increasing s . A similar behavior has already been found for the cuboctahedron, compare Ref. [100]. For the cases $s = 2$ and $s = 5/2$ we have calculated the lowest energy eigenvalues only in some M subspaces, including those subspaces that are relevant for the calculation of the plateau width. We have kept $m = 2000$ density matrix eigenstates for these calculations. The plateau is clearly visible.

In the following we focus on the $s = 5/2$ case and the comparison to previous results on this system. Fig. 13 shows the ground state energy as a function of the DMRG sweep and as a function of the kept density matrix eigenstates. One can see that even for 2000 kept states and more than 30 sweeps the result is not yet converged. A much larger number of states is needed to get convergence. Also, an extrapolation to $m \rightarrow \infty$ is not reliably possible because for that many more sweeps would have to be performed.

For $m = 2000$ we obtain the value $E_0^{\text{DMRG}} \approx -216.5 J$. This value can be compared with previous results. The DMRG result of Exler and Schnack for the ground state energy (with $m = 120$) is approximately $-211.1 J$ [39], a value that is much higher and thus much more imprecise than our result. The very recent result of Neuscamman and Chan using correlator product states in combination with variational Monte Carlo is $-216.3 J$ [103]. This demonstrates that modern DMRG is able to accurately estimate relative ground state energies, at least much more accurate than determined by other methods.

3.3 Evaluation of thermodynamic observables

Exact evaluation

For the sake of completeness we want to outline how basic observables can be evaluated both as function of temperature T and magnetic field B . We will assume that $[\tilde{H}, \tilde{S}_z] = 0$ for this part, so that the energy eigenvectors $|\nu\rangle$ can be chosen as simultaneous eigenvectors of \tilde{S}_z with eigenvalues $E_\nu(B)$ and M_ν . The dependence of $E_\nu(B)$ on B is simply given by the Zeeman term. If \tilde{H} and \tilde{S}_z do not commute the respective traces for the partition function and thermodynamic means have to be evaluated starting from their general definitions.

The partition function is defined as

$$Z(T, B) = \text{Tr} \left\{ e^{-\beta \tilde{H}} \right\} = \sum_{\nu} e^{-\beta E_{\nu}(B)} . \quad (31)$$

Then the magnetization and the susceptibility per molecule can be evaluated from the first and the second moment of \tilde{S}_z

$$\mathcal{M}(T, B) = -\frac{1}{Z} \text{Tr} \left\{ g\mu_B \tilde{S}_z e^{-\beta \tilde{H}} \right\} = -\frac{g\mu_B}{Z(T, B)} \sum_{\nu} M_{\nu} e^{-\beta E_{\nu}(B)} \quad (32)$$

$$\chi(T, B) = \frac{\partial \mathcal{M}(T, B)}{\partial B} = \frac{(g\mu_B)^2}{k_B T} \left\{ \frac{1}{Z} \sum_{\nu} M_{\nu}^2 e^{-\beta E_{\nu}(B)} - \left(\frac{1}{Z} \sum_{\nu} M_{\nu} e^{-\beta E_{\nu}(B)} \right)^2 \right\} \quad (33)$$

In a similar way the internal energy and the specific heat are evaluated from first and second moment of the Hamiltonian

$$U(T, B) = -\frac{1}{Z} \text{Tr} \left\{ \tilde{H} e^{-\beta \tilde{H}} \right\} = -\frac{1}{Z} \sum_{\nu} E_{\nu}(B) e^{-\beta E_{\nu}(B)} \quad (34)$$

$$\begin{aligned} C(T, B) &= \frac{\partial U(T, B)}{\partial T} \\ &= \frac{1}{k_B T^2} \left\{ \frac{1}{Z} \sum_{\nu} (E_{\nu}(B))^2 e^{-\beta E_{\nu}(B)} - \left(\frac{1}{Z} \sum_{\nu} E_{\nu}(B) e^{-\beta E_{\nu}(B)} \right)^2 \right\} . \end{aligned} \quad (35)$$

Approximate evaluation by means of the Finite-Temperature Lanczosmethod

For the evaluation of thermodynamic properties in the canonical ensemble the exact partition function Z depending on temperature T and magnetic field B is given by

$$Z(T, B) = \sum_{\nu} \langle \nu | e^{-\beta \tilde{H}} | \nu \rangle . \quad (36)$$

Here $\{|\nu\rangle\}$ denotes an arbitrary orthonormal basis of the respective Hilbert space. Following the ideas of Refs. [42, 43] the unknown matrix elements are approximated as

$$\langle \nu | e^{-\beta \tilde{H}} | \nu \rangle \approx \sum_{n=1}^{N_L} \langle \nu | n(\nu) \rangle e^{-\beta \epsilon_n^{(\nu)}} \langle n(\nu) | \nu \rangle . \quad (37)$$

For the evaluation of the right hand side of Eq. (37) $|\nu\rangle$ is taken as the initial vector of a Lanczos iteration. This iteration consists of N_L Lanczos steps, which span a respective Krylov space. As common for the Lanczos method the Hamiltonian is diagonalized in this Krylov space. This yields the N_L Lanczos eigenvectors $|n(\nu)\rangle$ as well as the associated Lanczos energy eigenvalues $\epsilon_n^{(\nu)}$. They are enumerated by $n = 1, \dots, N_L$. The notation $n(\nu)$ is chosen to remind one that the Lanczos eigenvectors $|n(\nu)\rangle$ belong to the Krylov space derived from the original state $|\nu\rangle$.

The number of Lanczos steps N_L is a parameter of the approximation that needs to be large enough to reach the extremal energy eigenvalues but should not be too large in order not to run into problems of numerical accuracy. $N_L \approx 100$ is a typical and good value.

In addition, the complete and thus very large sum over all $\nu = 1, \dots, \dim(\mathcal{H})$ states $|\nu\rangle$ is replaced by a summation over a subset of R random vectors. These vectors are truly random, they do not need to belong to any special basis set. Altogether this yields for the partition function

$$Z(T, B) \approx \frac{\dim(\mathcal{H})}{R} \sum_{\nu=1}^R \sum_{n=1}^{N_L} e^{-\beta \epsilon_n^{(\nu)}} |\langle n(\nu) | \nu \rangle|^2. \quad (38)$$

Although this already sketches the general idea, it will always improve the accuracy if symmetries are taken into account as in the following formulation

$$Z(T, B) \approx \sum_{\Gamma} \frac{\dim(\mathcal{H}(\Gamma))}{R_{\Gamma}} \sum_{\nu=1}^{R_{\Gamma}} \sum_{n=1}^{N_L} e^{-\beta \epsilon_n^{(\nu, \Gamma)}} |\langle n(\nu, \Gamma) | \nu, \Gamma \rangle|^2. \quad (39)$$

Here Γ labels the irreducible representations of the employed symmetry group. The full Hilbert space is decomposed into mutually orthogonal subspaces $\mathcal{H}(\Gamma)$.

An observable would then be calculated as

$$O(T, B) \approx \frac{1}{Z(T, B)} \sum_{\Gamma} \frac{\dim(\mathcal{H}(\Gamma))}{R_{\Gamma}} \sum_{\nu=1}^{R_{\Gamma}} \sum_{n=1}^{N_L} e^{-\beta \epsilon_n^{(\nu, \Gamma)}} \langle n(\nu, \Gamma) | Q | \nu, \Gamma \rangle \langle \nu, \Gamma | n(\nu, \Gamma) \rangle. \quad (40)$$

It was noted in Ref. [104] that this approximation of the observable $O(T, B)$ may contain large statistical fluctuations at low temperatures due to the randomness of the set of states $\{|\nu, \Gamma\rangle\}$. It was shown that this can largely be cured by assuming a symmetrized version of Eq. (40). For our investigations this is irrelevant.

Our very positive experience is that even for large problems the number of random starting vectors as well as the number of Lanczos steps can be chosen rather small, e.g. $R \approx 20$, $N_L \approx 100$ [44,45,72]. Figure 14 displays the zero-field differential susceptibility of the cuboctahedron with $s = 3/2$ as an example. One notices that the approximate result, that anyway deviates from the exact one only for $0.5 \leq k_B T / |J| \leq 3$, quickly approaches the exact curve with increasing number R of initial states. Already for $R = 20$ the approximation is practically indistinguishable from the exact one; an increase to $R = 100$ does not further improve this observable [44].

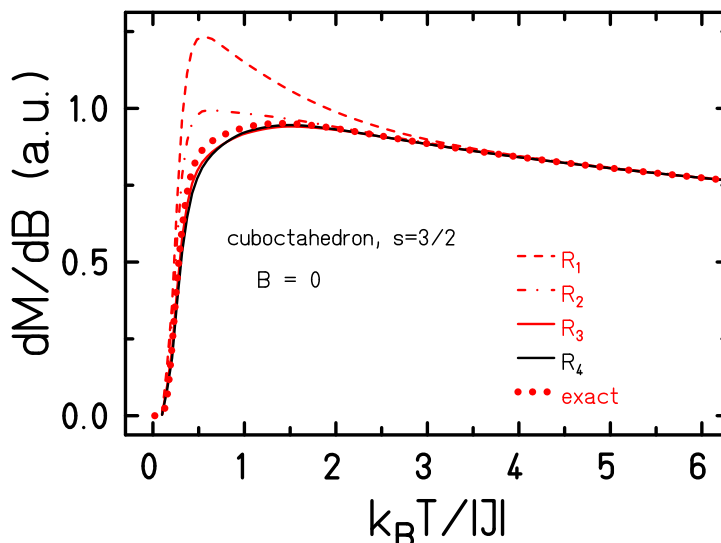


Fig. 14: Zero-field differential susceptibility of the cuboctahedron with $s = 3/2$. The various curves depict the investigated scenarios $R_1 = 1$, $R_2 = 5$, $R_3 = 20$, and $R_4 = 100$; $N_L = 100$. The exact dependence is given by the dots [44].

It is foreseeable that the method does not work optimally in very small subspaces or subspaces with large degeneracies of energy levels especially if the symmetry is not broken up into irreducible representations Γ . The underlying reason is given by the properties of the Lanczos method itself that fails to dissolve such degeneracies. The other case of small subspaces can be solved by including their energy eigenvalues and eigenstates exactly.

Another technical issue is given by the fact that the chosen random vectors $|\nu, \Gamma\rangle$ should be mutually orthogonal. Although one could orthonormalize the respective vectors, this is for practical purposes not really necessary. The reason is, that two vectors with random components are practically always orthogonal, because their scalar product is a sum over fluctuating terms that nearly vanishes especially in very large Hilbert spaces.

Since Lanczos iterations consist of matrix vector multiplications they can be parallelized by `openMP` directives. In our programs this is further accelerated by an analytical state coding and an evaluation of matrix elements of the Heisenberg Hamiltonian “on the fly” [105].

4 General properties of spectra

In the following I discuss some properties of the spectra of magnetic molecules with isotropic and antiferromagnetic interaction.

Non-bipartite spin rings

With the advent of magnetic molecules it appears to be possible to synthesize spin rings with an odd number of spins [106]. Although related to infinite spin rings and chains such systems have not been considered mainly since it does not really matter whether an infinite ring has an

odd or an even number of spins. In addition the sign rule of Marshall and Peierls [107] and the famous theorems of Lieb, Schultz, and Mattis [108, 109] provided valuable tools for the understanding of even rings which have the property to be bipartite and are thus non-frustrated. These theorems explain the degeneracy of the ground states in subspaces $\mathcal{H}(M)$ as well as their shift quantum number k or equivalently crystal momentum quantum number $p_k = 2\pi k/N$.

Nowadays exact diagonalization methods allow to evaluate eigenvalues and eigenvectors of \tilde{H} for small even and odd spin rings of various numbers N of spin sites and spin quantum numbers s where the interaction is given by antiferromagnetic nearest neighbor exchange [80–82, 110–112]. Although Marshall-Peierls sign rule and the theorems of Lieb, Schultz, and Mattis do not apply to non-bipartite rings, i.e. frustrated rings with odd N , it turns out that such rings nevertheless show astonishing regularities. Unifying the picture for even and odd N , we find for the ground state without exception [111–113]:

1. The ground state belongs to the subspace $\mathcal{H}(S)$ with the smallest possible total spin quantum number S ; this is either $S = 0$ for $N \cdot s$ integer, then the total magnetic quantum number M is also zero, or $S = 1/2$ for $N \cdot s$ half integer, then $M = \pm 1/2$.
2. If $N \cdot s$ is integer, then the ground state is non-degenerate.
3. If $N \cdot s$ is half integer, then the ground state is fourfold degenerate.
4. If s is integer or $N \cdot s$ even, then the shift quantum number is $k = 0$.
5. If s is half integer and $N \cdot s$ odd, then the shift quantum number turns out to be $k = N/2$.
6. If $N \cdot s$ is half integer, then $k = \lfloor (N+1)/4 \rfloor$ and $k = N - \lfloor (N+1)/4 \rfloor$ is found. $\lfloor (N+1)/4 \rfloor$ symbolizes the greatest integer less or equal to $(N+1)/4$.

In the case of $s = 1/2$ one knows the k -quantum numbers for all N via the Bethe ansatz [82, 110], and for spin $s = 1$ and even N the k quantum numbers are consistent with Ref. [81]. It appears that for the properties of the first excited state such rules do not hold in general, but only for “high enough” $N > 5$ [112]. Then, as can be anticipated from table 1, we can conjecture that

- if N is even, then the first excited state has $S = 1$ and is threefold degenerate, and
- if N is odd and the single particle spin is half-integer, then the first excited state has $S = 3/2$ and is eightfold degenerate, whereas
- if N is odd and the single particle spin is integer, then the first excited state has $S = 1$ and is sixfold degenerate.

Considering relative ground states in subspaces $\mathcal{H}(M)$ one also finds – for even as well as for odd N – that the shift quantum numbers k show a strikingly simple regularity for $N \neq 3$

$$k \equiv \pm(Ns - M) \left\lceil \frac{N}{2} \right\rceil \pmod{N}, \quad (41)$$

s	N									
	2	3	4	5	6	7	8	9	10	
$\frac{1}{2}$	1.5	0.5	1	0.747	0.934	0.816	0.913	0.844	0.903	$E_0/(NJ)$
	1	4	1	4	1	4	1	4	1	deg
	0	1/2	0	1/2	0	1/2	0	1/2	0	S
	1	1, 2	0	1, 4	3	2, 5	0	2, 7	5	k
$\frac{1}{2}$	4.0	3.0	2.0	2.236	1.369	2.098	1.045	1.722	0.846	$\Delta E/ J $
	3	4	3	2	3	8	3	8	3	deg
	1	3/2	1	1/2	1	3/2	1	3/2	1	S
	0	0	2	0	0	1, 6	4	3, 6	0	k
1	4	2	3	2.612	2.872	2.735	2.834	2.773	2.819	$E_0/(NJ)$
	1	1	1	1	1	1	1	1	1	deg
	0	0	0	0	0	0	0	0	0	S
	0	0	0	0	0	0	0	0	0	k
1	4.0	2.0	2.0	1.929	1.441	1.714	1.187	1.540	1.050	$\Delta E/ J $
	3	9	3	6	3	6	3	6	3	deg
	1	1	1	1	1	1	1	1	1	S
	1	0, 1, 2	2	2, 3	3	3, 4	4	4, 5	5	k
$\frac{3}{2}$	7.5	3.5	6	4.973	5.798	5.338	5.732	5.477	5.704	$E_0/(NJ)$
	1	4	1	4	1	4	1	4	1	deg
	0	1/2	0	1/2	0	1/2	0	1/2	0	S
	1	1, 2	0	1, 4	3	2, 5	0	2, 7	5	k
$\frac{3}{2}$	4.0	3.0	2.0	2.629	1.411	2.171	1.117	1.838	0.938	$\Delta E/ J $
	3	16	3	8	3	8	3	8	3	deg
	1	3/2	1	3/2	1	3/2	1	3/2	1	S
	0	0, 1, 2	2	2, 3	0	1, 6	4	3, 6	0	k
2	12	6	10	8.456	9.722	9.045	9.630	9.263	9.590	$E_0/(NJ)$
	1	1	1	1	1	1	1	1	1	deg
	0	0	0	0	0	0	0	0	0	S
	0	0	0	0	0	0	0	0	0	k
2	4.0	2.0	2.0	1.922	1.394	1.652	1.091	1.431	0.906	$\Delta E/ J $
	3	9	3	6	3	6	3	6	3	deg
	1	1	1	1	1	1	1	1	1	S
	1	0, 1, 2	2	2, 3	3	3, 4	4	4, 5	5	k
$\frac{5}{2}$	17.5	8.5	15	12.434	14.645	13.451	14.528	13.848	14.475	$E_0/(NJ)$
	1	4	1	4	1	4	1	4	1	deg
	0	1/2	0	1/2	0	1/2	0	1/2	0	S
	1	1, 2	0	1, 4	3	2, 5	0	2, 7	5	k

Table 1: Properties of ground and first excited state of AF Heisenberg rings for various N and s : ground state energy E_0 , gap ΔE , degeneracy deg , total spin S and shift quantum number k .

where $\lceil N/2 \rceil$ denotes the smallest integer greater than or equal to $N/2$ [113]. For $N = 3$ and $3s - 2 \geq |M| \geq 1$ one finds besides the ordinary k -quantum numbers given by (41) extraordinary k -quantum numbers, which supplement the ordinary ones to the complete set $\{k\} = \{0, 1, 2\}$.

For even N the k values form an alternating sequence $0, N/2, 0, N/2, \dots$ on descending from the magnon vacuum with $M = Ns$ as known from the sign-rule of Marshall and Peierls [107]. For odd N it happens that the ordinary k -numbers are repeated on descending from $M \leq Ns - 1$ to $M - 1$ iff N divides $[2(Ns - M) + 1]$.

Using the k -rule one can as well derive a rule for the relative ground state energies and for the respective S quantum numbers:

- For the relative ground state energies one finds that if the k -number is different in adjacent subspaces, $E_{\min}(S) < E_{\min}(S + 1)$ holds. If the k -number is the same, the energies could as well be the same.
- Therefore, if N (even or odd) does not divide $(2(Ns - M) + 1)[N/2]$, then any relative ground state in $\mathcal{H}(M)$ has the total spin quantum number $S = |M|$.
- This is always true for the absolute ground state which therefore has $S = 0$ for Ns integer and $S = 1/2$ for Ns half integer.

The k -rule (41) is founded in a mathematically rigorous way for N even [107–109], $N = 3$, $M = Ns$, $M = Ns - 1$, and $M = Ns - 2$ [113]. An asymptotic proof for large enough N can be provided for systems with an asymptotically finite excitation gap, i.e. systems with integer spin s for which the Haldane conjecture applies [114, 115]. In all other cases numerical evidence was collected and the k -rule as a conjecture still remains a challenge [113].

Rotational bands

For many spin systems with constant isotropic antiferromagnetic nearest neighbor Heisenberg exchange the minimal energies $E_{\min}(S)$ form a rotational band, i.e. depend approximately quadratically on the total spin quantum number S [101, 116, 117]

$$E_{\min}(S) \approx E_a - J \frac{D(N, s)}{N} S(S + 1). \quad (42)$$

The occurrence of a rotational band has been noted on several occasions for an even number of spins defining a ring structure, see, e.g., Ref. [117]. The minimal energies have been described as “following the Landé interval rule” [54–56, 58]. However, one finds that the same property also occurs for rings with an odd number of spins as well as for the various polytope configurations we have investigated, in particular for quantum spins positioned on the vertices of a tetrahedron, cube, octahedron, icosahedron, triangular prism, and an axially truncated icosahedron. Rotational modes have also been found in the context of finite square and triangular lattices of spin-1/2 Heisenberg antiferromagnets [118, 119].

There are several systems, like spin dimers, trimers, squares, tetrahedra, and octahedra which possess a strict rotational band since their Hamiltonian can be simplified by quadrature. As an example the Heisenberg square, i.e., a ring with $N = 4$ is presented. Because the Hamilton

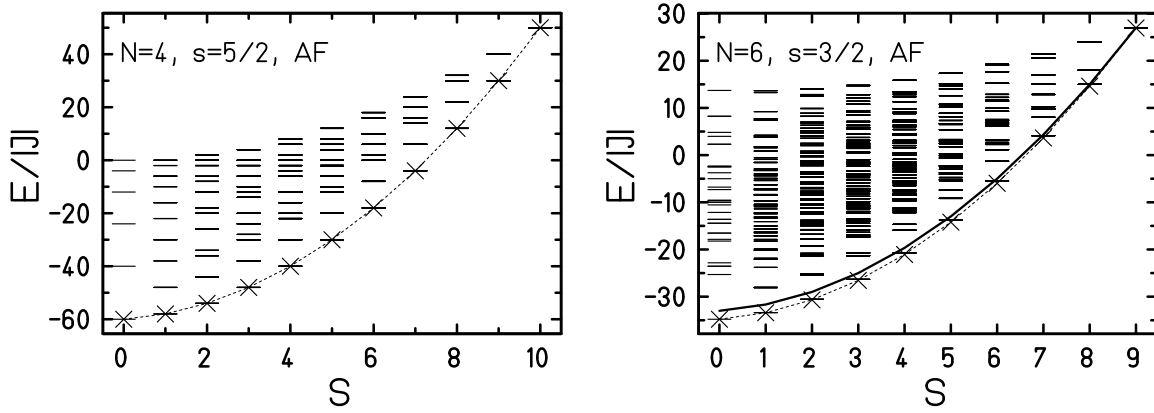


Fig. 15: Energy spectra of antiferromagnetically coupled Heisenberg spin rings (horizontal dashes). The crosses connected by the dashed line represent the fit to the rotational band according to (46), which matches both the lowest and the highest energies exactly. On the left the dashed line reproduces the exact rotational band, whereas on the right it only approximates it, but to high accuracy. The solid line on the right corresponds to the approximation Eq. (47).

operator (23) can be rewritten as

$$\tilde{H} = -J \left(\tilde{S}^2 - \tilde{S}_{13}^2 - \tilde{S}_{24}^2 \right), \quad (43)$$

$$\vec{\tilde{S}}_{13} = \vec{\tilde{s}}(1) + \vec{\tilde{s}}(3) \quad (44)$$

$$\vec{\tilde{S}}_{24} = \vec{\tilde{s}}(2) + \vec{\tilde{s}}(4), \quad (45)$$

with all spin operators \tilde{S}^2 , \tilde{S}_{13}^2 and \tilde{S}_{24}^2 commuting with each other and with \tilde{H} , one can directly obtain the complete set of eigenenergies, and these are characterized by the quantum numbers S , S_{13} and S_{24} . In particular, the lowest energy for a given total spin quantum number S occurs for the choice $S_{13} = S_{24} = 2s$

$$E_{min}(S) = -J [S(S+1) - 2 \cdot 2s(2s+1)] = E_0 - J S(S+1), \quad (46)$$

where $E_0 = 4s(2s+1)J$ is the exact ground state energy. The various energies $E_{min}(S)$ form a rigorous parabolic rotational band of excitation energies. Therefore, these energies coincide with a parabolic fit (crosses connected by the dashed line on the left of Fig. 15) passing through the antiferromagnetic ground state energy and the highest energy level, i.e., the ground state energy of the corresponding ferromagnetically coupled system.

It turns out that an accurate formula for the coefficient $D(N, s)$ of (46) can be developed using the sublattice structure of the spin array [116]. As an example we repeat the basic ideas for Heisenberg rings with an even number of spin sites [58]. Such rings are bipartite and can be decomposed into two sublattices, labeled A and B , with every second spin belonging to the same sublattice. The classical ground state (Néel state) is given by an alternating sequence of opposite spin directions. On each sublattice the spins are mutually parallel. Therefore, a quantum trial state, where the individual spins on each sublattice are coupled to their maximum values, $S_A = S_B = Ns/2$, could be expected to provide a reasonable approximation to the true

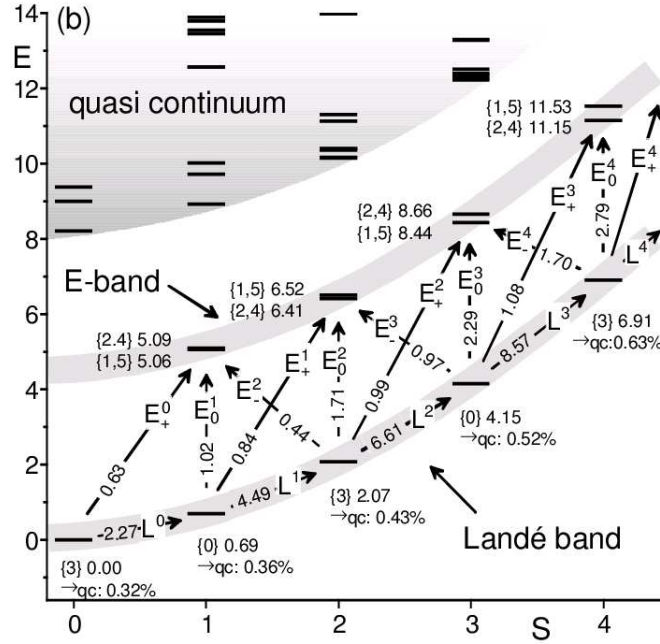


Fig. 16: The low-lying levels of a spin ring, $N = 6$ and $s = 5/2$ in this example, can be grouped into the lowest (Landé) band, the first excited (Excitation) band and the quasi-continuum (QC). For the spin levels of the L- and E-band k is given in brackets followed by the energy. Arrows indicate strong transitions from the L-band. Associated numbers give the total oscillator strength f_0 for these transitions [117].

ground state, especially if s assumes large values. For rings with even N the approximation to the respective minimal energies for each value of the total spin $\vec{S} = \vec{S}_A + \vec{S}_B$ is then [58]

$$E_{min}^{approx}(S) = -\frac{4J}{N} \left[S(S+1) - 2\frac{Ns}{2} \left(\frac{Ns}{2} + 1 \right) \right]. \quad (47)$$

This approximation exactly reproduces the energy of the highest energy eigenvalue, i.e., the ground state energy of the corresponding ferromagnetically coupled system ($S = Ns$). For all smaller S the approximate minimal energy $E_{min}^{approx}(S)$ is bounded from below by the true one (Rayleigh-Ritz variational principle). The solid curve displays this behavior for the example of $N = 6$, $s = 3/2$ on the right of Fig. 15. The coefficient “4” in Eq. (47) is the classical value, i.e. for each fixed even N the coefficient $D(N, s)$ approaches 4 with increasing s [116, 120].

The approximate spectrum, (47), is similar to that of two spins, \vec{S}_A and \vec{S}_B , each of spin quantum number $Ns/2$, that are coupled by an effective interaction of strength $4J/N$. Therefore, one can equally well say that the approximate rotational band considered in (47) is associated with an effective Hamilton operator

$$\tilde{H}^{approx} = -\frac{4J}{N} \left[\vec{S}^2 - \vec{S}_A^2 - \vec{S}_B^2 \right], \quad (48)$$

where the two sublattice spins, \vec{S}_A, \vec{S}_B , assume their maximal value $S_A = S_B = Ns/2$. Hamiltonian (48) is also known as Hamiltonian of the Lieb-Mattis model which describes a system

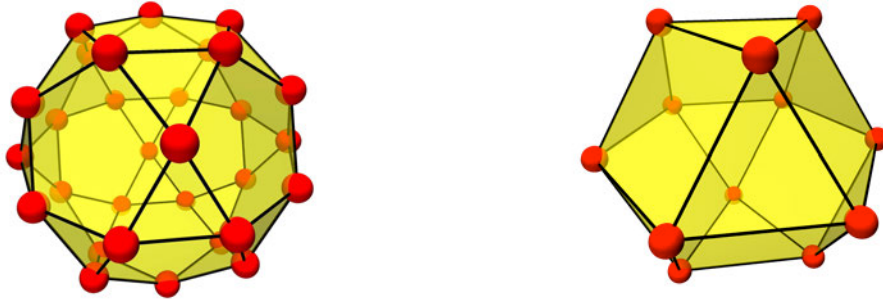


Fig. 17: Structure of the icosidodecahedron (left) and the cuboctahedron (right).

where each spin of one sublattice interacts with every spin of the other sublattice with equal strength [109, 121].

It is worth noting that this Hamiltonian reproduces more than the lowest levels in each subspace $\mathcal{H}(S)$. At least for bipartite systems also a second band is accurately reproduced as well as the gap to the quasi-continuum above, compare Fig. 16. This property is very useful since the approximate Hamiltonian allows the computation of several observables without diagonalizing the full Hamiltonian.

Magnetization jumps

Although the spectra of many magnetic molecules possess an approximate rotational band of minimal energies $E_{min}(S)$ and although in the classical limit, where the single-spin quantum number s goes to infinity, the function $E_{min}(S)$ is even an exact parabola if the system has coplanar ground states [122], one finds that for certain coupling topologies, including the cuboctahedron and the icosidodecahedron (see Fig. 17), that this rule is violated, e.g. for high total spins [102, 123, 124]. More precisely, for the icosidodecahedron the last four points of the graph of E_{min} versus S , i. e. the points with $S = S_{max}$ to $S = S_{max} - 3$, lie on a straight line

$$E_{min}(S) = 60Js^2 - 6Js(30s - S) . \quad (49)$$

An analogous statement holds for the last three points of the corresponding graph for the cuboctahedron. These findings are based on numerical calculations of the minimal energies for several s both for the icosidodecahedron as well as for the cuboctahedron. For both these and other systems a rigorous proof of the high spin anomaly can be given [102, 125].

The idea of the proof can be summarized as follows: A necessary condition for the anomaly is certainly that the minimal energy in the one-magnon space is degenerate. Therefore, localized one-magnon states can be constructed which are also of minimal energy. When placing a second localized magnon on the spin array, there will be a chance that it does not interact with the first one if a large enough separation can be achieved. This new two-magnon state is likely the state of minimal energy in the two-magnon Hilbert space because for antiferromagnetic interaction two-magnon bound states do not exist. This procedure can be continued until no further independent magnon can be placed on the spin array. In a sense the system behaves as if it consists of non-interacting bosons which, up to a limiting number, can condense into

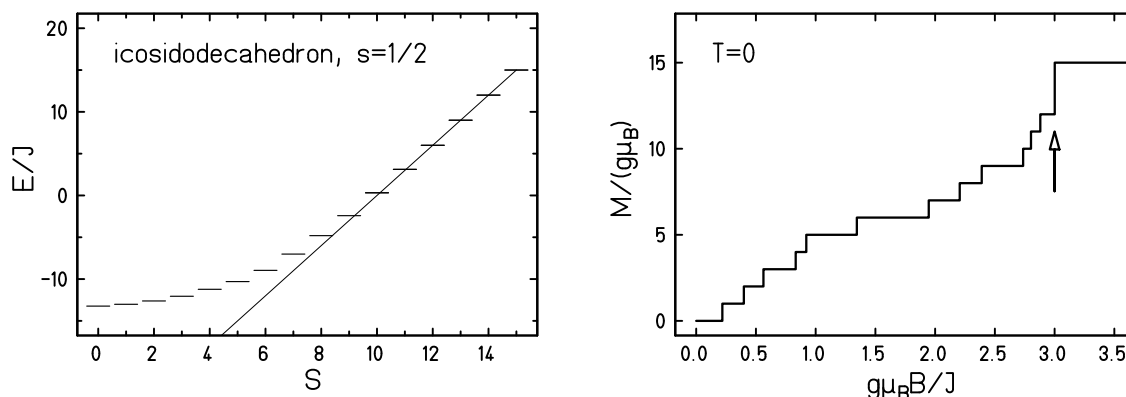


Fig. 18: Icosidodecahedron: Left: minimal energy levels $E_{\min}(S)$ as a function of total spin S . Right: magnetization curve at $T = 0$ [102].

a single-particle ground state. In more mathematical terms: To prove the high-spin anomaly one first shows an inequality which says that all points $(S, E_{\min}(S))$ lie above or on the line connecting the last two points. For specific systems as those mentioned above what remains to be done is to construct particular states that exactly assume the values of E_{\min} corresponding to the points on the bounding line, then these states are automatically of minimal energy.

The observed anomaly (linear instead of parabolic dependence) results in a corresponding jump of the magnetization curve \mathcal{M} vs. B , see Fig. 18. In contrast, for systems which obey the Landé interval rule the magnetization curve at very low temperatures is a staircase with equal steps up to the highest magnetization. The anomaly could indeed be observed in magnetization measurements of the Keplerate $\{\text{Mo}_{72}\text{Fe}_{30}\}$. Unfortunately, the magnetization measurements [66, 101] performed so far suffer from too high temperatures which smear out the anomaly.

Nevertheless, it may be possible to observe truly giant magnetization jumps in certain two-dimensional spin systems which possess a suitable coupling (e.g. Kagomé) [123, 126]. In such systems the magnetization jump can be of the same order as the number of spins, i.e. the jump remains finite – or in other words is macroscopic – in the thermodynamic limit $N \rightarrow \infty$. Thus, this effect is a true macroscopic quantum effect. It is also related to so-called flat bands of energy eigenvalues [124, 127–131].

5 Magnetocalorics

The use of magnetic molecules for sub-Kelvin cooling is one of the later ideas in the field of molecular magnetism. The mean (internal) energy, the magnetization and the magnetic field are thermodynamic observables just as pressure and volume. Therefore, we can design thermodynamic processes which work with magnetic materials as a medium. This has of course already been done for a long time. The most prominent application is magnetization cooling which is mainly used to reach sub-Kelvin temperatures [132]. The first observation of sub-Kelvin temperatures is a nice example of how short an article can be to win the Nobel prize (Giaque, Chemistry, 1949). Nowadays magnetization cooling is used in ordinary refrigerators.

Attainment of Temperatures Below 1° Absolute by Demagnetization of $\text{Gd}_2(\text{SO}_4)_3 \cdot 8\text{H}_2\text{O}$

We have recently carried out some preliminary experiments on the adiabatic demagnetization of $\text{Gd}_2(\text{SO}_4)_3 \cdot 8\text{H}_2\text{O}$ at the temperatures of liquid helium. As previously predicted by one of us, a large fractional lowering of the absolute temperature was obtained.

An iron-free solenoid producing a field of about 8000 gauss was used for all the measurements. The amount of $\text{Gd}_2(\text{SO}_4)_3 \cdot 8\text{H}_2\text{O}$ was 61 g. The observations were checked by many repetitions of the cooling. The temperatures were measured by means of the inductance of a coil surrounding the gadolinium sulfate. The coil was immersed in liquid helium and isolated from the gadolinium by means of an evacuated space. The thermometer was in excellent agreement with the temperature of liquid helium as indicated by its vapor pressure down to 1.5°K.

On March 19, starting at a temperature of about 3.4°K, the material cooled to 0.53°K. On April 8, starting at about 2°, a temperature of 0.34°K was reached. On April 9, starting at about 1.5°, a temperature of 0.25°K was attained.

It is apparent that it will be possible to obtain much lower temperatures, especially when successive demagnetizations are utilized.

W. F. GIAUQUE
D. P. MACDOUGALL

Department of Chemistry,
University of California,
Berkeley, California,
April 12, 1933.

Fig. 19: *The first observation of sub-Kelvin temperatures [132] is a nice example of how short an article can be to win the Nobel prize (Giauque, Chemistry, 1949).*

In early magnetocaloric experiments simple refrigerants like paramagnetic salts have been used. We will therefore consider such examples first. For a paramagnet the Hamiltonian consists of the Zeeman term only. We then obtain for the partition function

$$Z(T, B, N) = \left(\frac{\sinh[\beta g \mu_B B (s + 1/2)]}{\sinh[\beta g \mu_B B / 2]} \right)^N. \quad (50)$$

Then the magnetization is

$$\mathcal{M}(T, B, N) = N g \mu_B \left((s + 1/2) \coth[\beta g \mu_B B (s + 1/2)] - 1/2 \sinh[\beta g \mu_B B / 2] \right), \quad (51)$$

and the entropy reads

$$S(T, B, N) = N k_B \ln \left(\frac{\sinh[\beta g \mu_B B (s + 1/2)]}{\sinh[\beta g \mu_B B / 2]} \right) - k_B \beta B \mathcal{M}(T, B, N). \quad (52)$$

Besides their statistical definition both quantities follow from the general thermodynamic relationship

$$dF = \left(\frac{\partial F}{\partial T} \right)_B dT + \left(\frac{\partial F}{\partial B} \right)_T dB = -SdT - \mathcal{M}dB, \quad (53)$$

where $F(T, B, N) = -k_B T \ln[Z(T, B, N)]$.

Looking at Eq. (50) it is obvious that all thermodynamic observables for a paramagnet depend on temperature and field via the combination B/T , and so does the entropy. Therefore, an adiabatic demagnetization ($S = \text{const}$) means that the ratio B/T has to remain constant, and thus temperature shrinks linearly with field, i.e.

$$\left(\frac{\partial T}{\partial B} \right)_S^{\text{para}} = \frac{T}{B}. \quad (54)$$

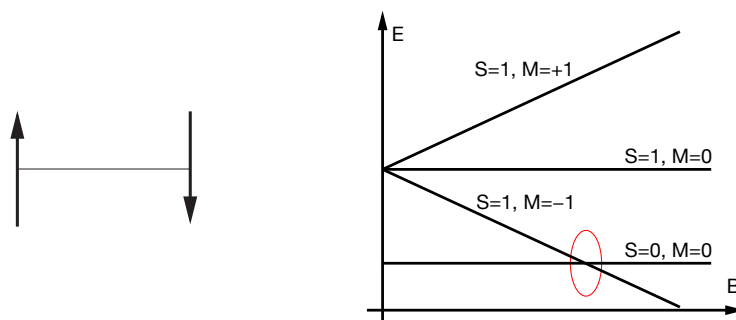


Fig. 20: Left: Sketch of an antiferromagnetically coupled spin dimer. Right: Dependence of the energy levels on magnetic field for an antiferromagnetically coupled spin-1/2 dimer. At the critical field B_c the lowest triplet level crosses the ground state level with $S = 0$.

This situation changes completely for an interacting spin system. Depending on the interactions the adiabatic cooling rate $\frac{\partial T}{\partial B}$ can be smaller or bigger than the paramagnetic one (54) and even change sign, i.e. one would observe heating during demagnetization. It is nowadays understood that the cooling rate acquires extreme values close to phase transitions due to the excess entropy associated with such processes [133–136].

In the following this statement will be made clear by discussing the example of a simple antiferromagnetically coupled spin-1/2 dimer, shown on the left of Fig. 20. In a magnetic field such a system experiences a “quantum phase transition” if the lowest triplet level crosses the original ground state with $S = 0$, see Fig. 20 on the right. Although one would hesitate to call such an ordinary ground state level crossing quantum phase transition it nevertheless is one. At $T = 0$ the magnetization $\mathcal{M}(T = 0, B)$ is a non-analytic function of the magnetic field B . At the critical field B_c where the levels cross the magnetization exhibits a step.

In addition the entropy, which at $T = 0$ is zero for the non-degenerate ground state acquires a finite value at the critical field B_c due to the degeneracy of the crossing levels. This enhancement remains present even at temperatures $T > 0$, it is shown on the left of Fig. 21. In addition the heat capacity varies strongly around the critical field as is shown on the right of Fig. 21.

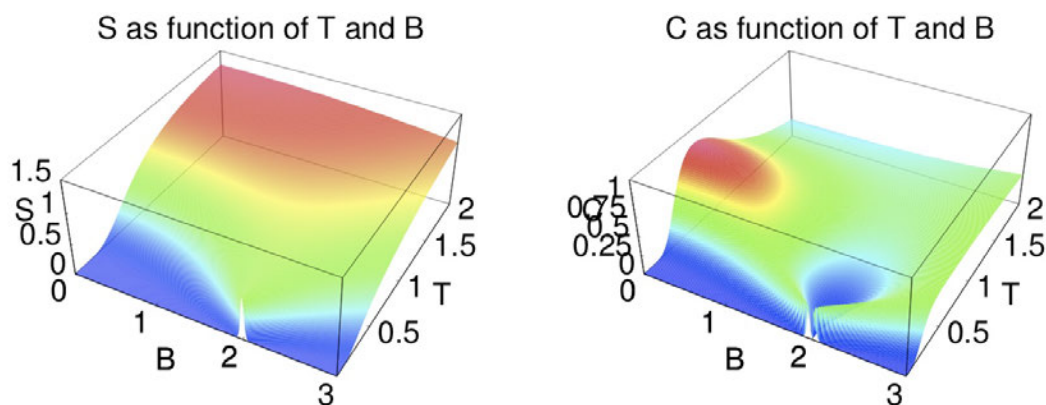


Fig. 21: Left: Entropy of the dimer (Fig. 20) as function of B and T . Right: Heat capacity of the dimer as function of B and T .

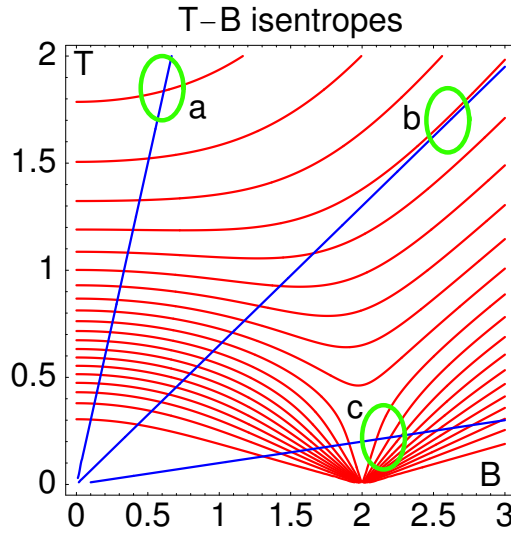


Fig. 22: Isentropes of the spin dimer. The straight lines show the behavior of a paramagnet for comparison. B is along the x -axis, T along the y -axis.

The behavior of the entropy as well as the heat capacity explains how the adiabatic cooling rate

$$\left(\frac{\partial T}{\partial B}\right)_S = -T \frac{\left(\frac{\partial S}{\partial B}\right)_T}{C} \quad (55)$$

depends on field and temperature. Figure 22 shows the isentropes of the antiferromagnetically coupled dimer both as function of field B and temperature T . The straight lines show the behavior of a paramagnet for comparison. Three regions are highlighted.

- a: For low fields and high temperatures $\frac{\partial T}{\partial B}$ is smaller than for the paramagnet.
- b: For high fields and high temperatures the interacting system assumes the paramagnetic limit, i.e. $\frac{\partial T}{\partial B}$ is the same in both systems.
- c: For low temperatures and fields just above the critical field $\frac{\partial T}{\partial B}$ is much bigger than the cooling rate of the paramagnet.
- Not highlighted but nevertheless very exciting is the region at low temperature just below the critical field where the “cooling” rate $\frac{\partial T}{\partial B}$ has the opposite sign, i.e. upon demagnetizing the system heats up and upon magnetizing the system cools down.

The rate $\frac{\partial T}{\partial B}$, Eq. (55), depends directly on the derivative of the entropy with respect to the magnetic field. Therefore, it is clear that the effect will be enhanced if a high degeneracy can be obtained at some critical field. This is indeed possible in several frustrated materials where giant magnetization jumps to saturation are observed [123, 124, 135–137].

As a final example I would like to discuss the $M=\text{Cu}$ and $M=\text{Ni}$ members of the family of Gd_4M_8 molecules which were synthesized quite recently [13]. The eigenvalues of the respective spin Hamiltonians could be determined numerically exactly for the case of Gd_4Cu_8 and by means of the Finite-Temperature Lanczos Method for Gd_4Ni_8 [45].

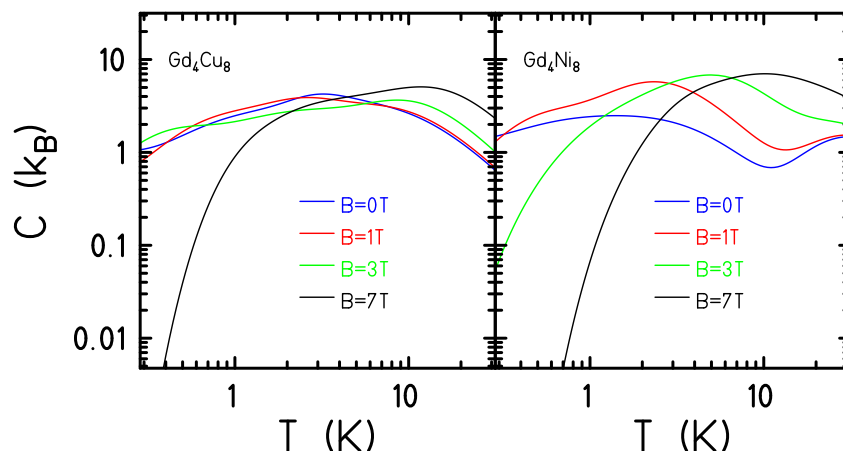


Fig. 23: Theoretical heat capacity per molecule for Gd_4Cu_8 (left) and Gd_4Ni_8 (right) at various magnetic fields.

For Gd_4Cu_8 the model Hamiltonian (2) includes the following parameters: $J_{GdGd} = -0.1 \text{ cm}^{-1}$, $J_{GdCu} = +0.9 \text{ cm}^{-1}$, $J_{CuCu} = -8.0 \text{ cm}^{-1}$. The spectroscopic splitting factor was taken as $g = 2.0$. For Gd_4Ni_8 the model parameters were chosen as: $J_{GdGd} = -0.1 \text{ cm}^{-1}$, $J_{GdNi} = +0.17 \text{ cm}^{-1}$, $J_{NiNi} = +12.0 \text{ cm}^{-1}$. Again we took $g = 2.0$. In order to provide a feeling for the energy scale of inverse centimeters we would like to mention that one inverse centimeter (wave number) corresponds to about 1.44 K in thermal energy. A coupling of about a tenth of an inverse centimeter or Kelvin is typical for interactions between gadolinium ions. In the following all other interactions or corrections such as temperature independent paramagnetism, different g factors for different ions or a possible single-ion anisotropy in the case of nickel have been neglected. Despite these approximations all theoretical curves agree nicely with the experimental ones published in Ref. [13].

Figure 23 displays the theoretical heat capacity per molecule for Gd_4Cu_8 (left) and Gd_4Ni_8 (right) at various magnetic fields. The behavior is for both compounds qualitatively similar.

The isothermal magnetic entropy change, shown in 24, turns out to be very different; it is much

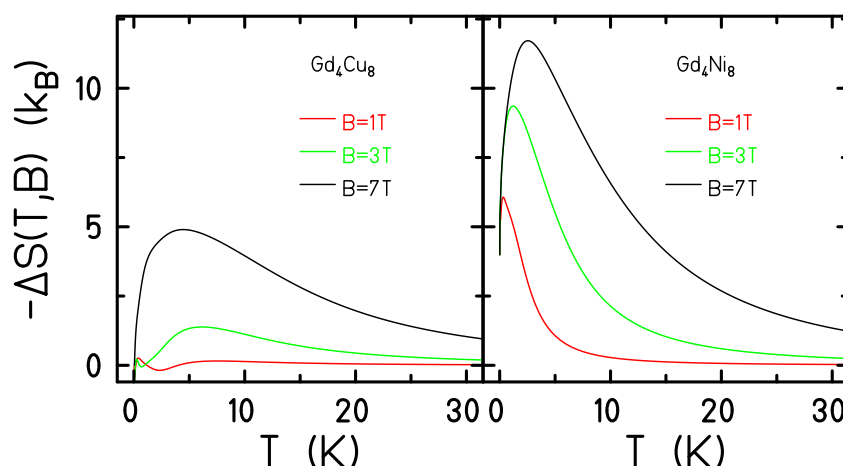


Fig. 24: Theoretical isothermal entropy change per molecule for Gd_4Cu_8 (left) and Gd_4Ni_8 (right) for various field differences: $-\Delta S(T, B) = -[S(T, B) - S(T, 0)]$.

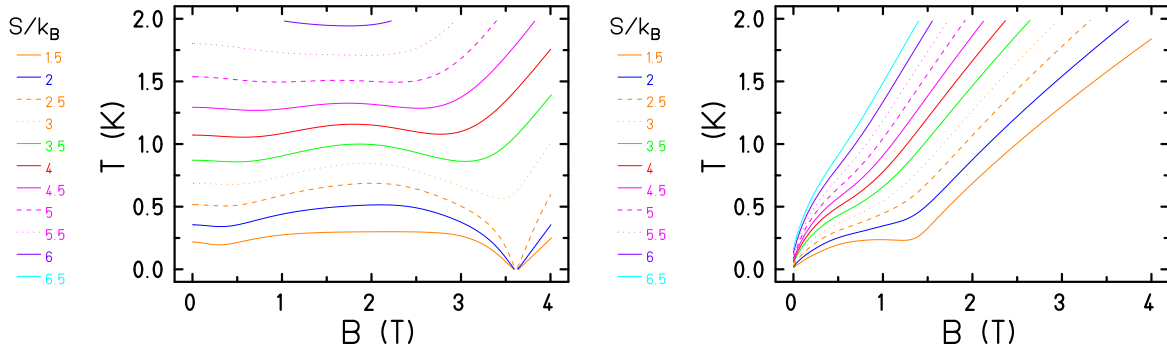


Fig. 25: Theoretical isentropes for Gd_4Cu_8 (left) and for Gd_4Ni_8 (right) [13, 45].

larger for Gd_4Ni_8 . The reason is that for Gd_4Ni_8 the low-lying multiplets belong to large total spin quantum numbers which leads to larger entropies at low temperatures. This is made even clearer by Fig. 25 displaying the isentropes as function of both temperature and magnetic field: Gd_4Cu_8 (left) possesses a non-degenerate $S_t = 0$ ground state that is separated from a triplet and a quintet, whereas Gd_4Ni_8 (right) has a ground state with $S_t = 22$. In the latter case all isentropes with $S \leq k_B \log(45)$ run into absolute zero, which is clearly visible in Fig. 25. On the contrary, since Gd_4Cu_8 possesses a non-degenerate $S_t = 0$ ground state all isentropes approach temperatures $T > 0$ when B goes to zero.

Although this behavior suggests that Gd_4Ni_8 should be a very good refrigerant, this does not need to be the case. At sub-Kelvin temperatures dipolar interactions become very important. They prevent a closer approach of $T = 0$ [138]. Dipolar interactions could be tamed by molecules that possess an $S_t = 0$ ground state, but a non-degenerate ground state would not be helpful due to its vanishing entropy. Therefore, we suggest to investigate molecules which have a degenerate – the more the better – ground state with $S_t = 0$. A ground state degeneracy can be induced by frustration, thus a tetrahedron with antiferromagnetic coupling would be a first candidate [45].

Summarizing, one can say that low-dimensional frustrated spin systems and especially magnetic molecules are substances with an interesting magnetocaloric behavior and may turn out to be useful new refrigerants for special applications.

Acknowledgment

I would like to thank my colleagues Klaus Bärwinkel, Mirko Brüger, Matthias Exler, Peter Hage, Christian Heesing, Detlef Mentrup, Andrei Postnikov, Johannes Richter, Heinz-Jürgen Schmidt, Roman Schnalle, Jörg Ummethum, Oliver Wendland as well as Marshall Luban and Christian Schröder for the fruitful collaboration on theoretical issues which produced many of the discussed results. I would also like to thank all my collaborators from chemistry and experimental physics with whom to work was and is a real pleasure.

This work was supported by the Deutsche Forschungsgemeinschaft (e.g. Grants No. SCHN 615/5-1, SCHN 615/8-1, SCHN 615/15-1, FOR 945).

References

- [1] J. Schnack: *Quantum Magnetism*, pp. 155–194 in *Lecture Notes in Physics*, Vol. 645 (Springer, Berlin, Heidelberg, 2004)
- [2] J. Schnack: *Quantum Theory of Molecular Magnetism*, pp.F1.1–F1.36 in *Magnetism goes Nano*, 36th IFF Spring School (Forschungszentrum Jülich, 2005)
- [3] R. Sessoli, D. Gatteschi, A. Caneschi, and M. A. Novak, *Nature* **365**, 141 (1993)
- [4] D. Gatteschi, A. Caneschi, L. Pardi, and R. Sessoli, *Science* **265**, 1054 (1994)
- [5] D. Gatteschi, *Adv. Mater.* **6**, 635 (1994)
- [6] A. Müller, F. Peters, M. Pope, and D. Gatteschi, *Chem. Rev.* **98**, 239 (1998)
- [7] A.L. Dearden, S. Parsons, and R.E.P. Winpenny, *Angew. Chem. Int. Ed.* **40**, 151 (2001)
- [8] P. King, T.C. Stamatatos, K.A. Abboud, and G. Christou, *Angew. Chem. Int. Ed.* **45**, 7379 (2006)
- [9] A.M. Ako, V. Mereacre, R. Clerac, I.J. Hewitt, Y. Lan, C.E. Anson, and A.K. Powell, *Dalton Trans.* **36**, 5245 (2007)
- [10] I.S. Tidmarsh, L.J. Batchelor, E. Scales, R.H. Laye, L. Sorace, A. Caneschi, J. Schnack, and E.J.L. McInnes, *Dalton Trans.* **38**, 9402 (2009)
- [11] K. Graham, F.J. Douglas, J.S. Mathieson, S.A. Moggach, J. Schnack, and M. Murrie, *Dalton Trans.* **40**, 12 271 (2011)
- [12] T. Glaser, *Chem. Commun.* **47**, 116 (2011)
- [13] T.N. Hooper, J. Schnack, S. Piligkos, M. Evangelisti, and E.K. Brechin, *Angew. Chem. Int. Ed.* **51**, 4633 (2012)
- [14] A. Müller, S. Sarkar, S.Q.N. Shah, H. Bögge, M. Schmidtman, S. Sarkar, P. Kögerler, B. Hauptfleisch, A. Trautwein, and V. Schünemann, *Angew. Chem. Int. Ed.* **38**, 3238 (1999)
- [15] E. Coronado, P. Delhaes, D. Gatteschi, and J. Miller (Eds.): *Localized and Itinerant Molecular Magnetism: From Molecular Assemblies to the Devices*, NATO ASI, Series E, Vol. 321 (Kluwer Academic, Dordrecht, 1996)
- [16] M.N. Leuenberger and D. Loss, *Nature* **410**, 789 (2001)
- [17] A. Ardavan, O. Rival, J.J.L. Morton, S.J. Blundell, A.M. Tyryshkin, G.A. Timco, and R.E.P. Winpenny, *Phys. Rev. Lett.* **98**, 057201 (2007)

- [18] A. Candini, G. Lorusso, F. Troiani, A. Ghirri, S. Carretta, P. Santini, G. Amoretti, C. Muryn, F. Tuna, G. Timco, E.J.L. McInnes, R.E.P. Winpenny, W. Wernsdorfer, and M. Affronte, *Phys. Rev. Lett.* **104**, 037203 (2010)
- [19] P. Santini, S. Carretta, F. Troiani, and G. Amoretti, *Phys. Rev. Lett.* **107**, 230502 (2011)
- [20] P. Gülich, A. Hauser, and H. Spiering, *Angew. Chem.* **106**, 2109 (1994)
- [21] Y.-Z. Zheng, M. Evangelisti, and R.E.P. Winpenny, *Chem. Sci.* **2**, 99 (2011)
- [22] J.W. Sharples, Y.-Z. Zheng, F. Tuna, E.J.L. McInnes, and D. Collison, *Chem. Commun.* **47**, 7650 (2011)
- [23] X. Chen, Y.-S. Fu, S.-H. Ji, T. Zhang, P. Cheng, X.-C. Ma, X.-L. Zou, W.-H. Duan, J.-F. Jia, and Q.-K. Xue, *Phys. Rev. Lett.* **101**, 197208 (2008)
- [24] M. Mannini, F. Pineider, P. Saintavrit, C. Danieli, E. Otero, C. Sciancalepore, A.M. Talarico, M.-A. Arrio, A. Cornia, D. Gatteschi, and R. Sessoli, *Nat. Mater.* **8**, 194 (2009)
- [25] M. Bernien, J. Miguel, C. Weis, M.E. Ali, J. Kurde, B. Krumme, P.M. Panchmatia, B. Sanyal, M. Piantek, P. Srivastava, K. Baberschke, P.M. Oppeneer, O. Eriksson, W. Kuch, and H. Wende, *Phys. Rev. Lett.* **102**, 047202 (2009)
- [26] M. Höck and J. Schnack, *Phys. Rev. B* **87**, 184408 (2013)
- [27] W. Heisenberg, *Z. f. Phys.* **49**, 619 (1928)
- [28] D. Gatteschi, R. Sessoli, and J. Villain: *Molecular Nanomagnets* (Oxford University Press, Oxford, 2006)
- [29] A. Bencini and D. Gatteschi:
Electron paramagnetic resonance of exchange coupled systems
(Springer, Berlin, Heidelberg, 1990)
- [30] C. Delfs, D. Gatteschi, L. Pardi, R. Sessoli, K. Wieghardt, and D. Hanke, *Inorg. Chem.* **32**, 3099 (1993)
- [31] D. Gatteschi and L. Pardi, *Gazz. Chim. Ital.* **123**, 231 (1993)
- [32] J.J. Borrás-Almenar, J.M. Clemente-Juan, E. Coronado, and B.S. Tsukerblat, *Inorg. Chem.* **38**, 6081 (1999)
- [33] O. Waldmann, *Phys. Rev. B* **61**, 6138 (2000)
- [34] B. Tsukerblat, *Inorg. Chim. Acta* **361**, 3746 (2008)
- [35] A.S. Boyarchenkov, I.G. Bostrem, and A.S. Ovchinnikov, *Phys. Rev. B* **76**, 224410 (2007)

- [36] R. Schnalle and J. Schnack, *Phys. Rev. B* **79**, 104419 (2009)
- [37] R. Schnalle and J. Schnack, *Int. Rev. Phys. Chem.* **29**, 403 (2010)
- [38] S.R. White, *Phys. Rev. B* **48**, 10345 (1993)
- [39] M. Exler and J. Schnack, *Phys. Rev. B* **67**, 094440 (2003)
- [40] U. Schollwöck, *Rev. Mod. Phys.* **77**, 259 (2005)
- [41] J. Ummethum, J. Nehr Korn, S. Mukherjee, N.B. Ivanov, S. Stuiber, T. Strässle, P.L.W. Tregenna-Piggott, H. Mutka, G. Christou, O. Waldmann, and J. Schnack, *Phys. Rev. B* **86**, 104403 (2012)
- [42] J. Jaklic and P. Prelovsek, *Phys. Rev. B* **49**, 5065 (1994)
- [43] J. Jaklic and P. Prelovsek, *Adv. Phys.* **49**, 1 (2000)
- [44] J. Schnack and O. Wendland, *Eur. Phys. J. B* **78**, 535 (2010)
- [45] J. Schnack and C. Heesing, *Eur. Phys. J. B* **86**, 46 (2013)
- [46] T. Lis, *Acta Chrytallogr. B* **36**, 2042 (1980)
- [47] L. Thomas, F. Lioni, R. Ballou, D. Gatteschi, R. Sessoli, and B. Barbara, *Nature* **383**, 145 (1996)
- [48] Z.H. Jang, A. Lascialfari, F. Borsa, and D. Gatteschi, *Phys. Rev. Lett.* **84**, 2977 (2000)
- [49] Y. Furukawa, K. Watanabe, K. Kumagai, F. Borsa, and D. Gatteschi, *Phys. Rev. B* **64**, 104401 (2001)
- [50] L. Thomas and B. Barbara, *J. Low Temp. Phys.* **113**, 1055 (1998)
- [51] N. Regnault, T. Jolicœur, R. Sessoli, D. Gatteschi, and M. Verdaguer, *Phys. Rev. B* **66**, 054409 (2002)
- [52] G. Chaboussant, A. Sieber, S. Ochsenbein, H.-U. Güdel, M. Murrie, A. Honecker, N. Fukushima, and B. Normand, *Phys. Rev. B* **70**, 104422 (2004)
- [53] C. A. Christmas, H.L. Tsai, L. Pardi, J.M. Kesselman, P.K. Gantzel, R.K. Chadha, D. Gatteschi, D.F. Harvey, and D.N. Hendrickson, *J. Am. Chem. Soc.* **115**, 12483 (1993)
- [54] K.L. Taft, C.D. Delfs, G.C. Papaefthymiou, S. Foner, D. Gatteschi, and S.J. Lippard, *J. Am. Chem. Soc.* **116**, 823 (1994)
- [55] A. Lascialfari, D. Gatteschi, F. Borsa, and A. Cornia, *Phys. Rev. B* **55**, 14341 (1997)
- [56] A. Lascialfari, D. Gatteschi, F. Borsa, and A. Cornia, *Phys. Rev. B* **56**, 8434 (1997)

- [57] M.-H. Julien, Z. Jang, A. Lascialfari, F. Borsa, M. Horvatić, A. Caneschi, and D. Gatteschi, *Phys. Rev. Lett.* **83**, 227 (1999)
- [58] G. Abbati, A. Caneschi, A. Cornia, A. Fabretti, and D. Gatteschi, *Inorg. Chim. Acta* **297**, 291 (2000)
- [59] O. Waldmann, R. Koch, S. Schromm, J. Schüle, P. Müller, I. Bernt, R.W. Saalfrank, F. Hampel, and E. Baltes, *Inorg. Chem.* **40**, 2986 (2001)
- [60] F. Meier and D. Loss, *Phys. Rev. Lett.* **86**, 5373 (2001)
- [61] O. Waldmann, J. Schüle, R. Koch, P. Müller, I. Bernt, R.W. Saalfrank, H.P. Andres, H.U. Güdel, and P. Allenspach, *Inorg. Chem.* **38**, 5879 (1999)
- [62] R.W. Saalfrank, I. Bernt, M.M. Chowdhry, F. Hampel, and G.B.M. Vaughan, *Chem. Eur. J.* **7**, 2765 (2001)
- [63] F. Larsen, E. McInnes, H. El Mkami, G. Rajaraman, E. Rentschler, A. Smith, G. Smith, V. Boote, M. Jennings, G. Timco, and R. Winpenny, *Angew. Chem. Int. Edit.* **42**, 101 (2003)
- [64] F. Larsen, J. Overgaard, S. Parsons, E. Rentschler, A. Smith, G. Timco, and R. Winpenny, *Angew. Chem. Int. Edit.* **42**, 5978 (2003)
- [65] U. Kortz, A. Müller, J. van Slageren, J. Schnack, N. S. Dalal, and M. Dressel, *Coord. Chem. Rev.* **253**, 2315 (2009)
- [66] A. Müller, M. Luban, C. Schröder, R. Modler, P. Kögerler, M. Axenovich, J. Schnack, P.C. Canfield, S. Bud'ko, and N. Harrison, *Chem. Phys. Chem.* **2**, 517 (2001)
- [67] M. Axenovich and M. Luban, *Phys. Rev. B* **63**, 100407 (2001)
- [68] G. Seeber, P. Kögerler, B.M. Kariuki, and L. Cronin, *Chem. Commun.* pp. 1580–1581 (2004)
- [69] J. Schnack, H. Nojiri, P. Kögerler, G.J.T. Cooper, and L. Cronin, *Phys. Rev. B* **70**, 174420 (2004)
- [70] N.B. Ivanov, J. Schnack, R. Schnalle, J. Richter, P. Kögerler, G. N. Newton, L. Cronin, Y. Oshima, and H. Nojiri, *Phys. Rev. Lett.* **105**, 037206 (2010)
- [71] V. Hoeke, M. Heidemeier, E. Krickemeyer, A. Stammler, H. Bögge, J. Schnack, A. Postnikov, and T. Glaser, *Inorg. Chem.* **51**, 10929 (2012)
- [72] Y. Zheng, Q.-C. Zhang, L.-S. Long, R.-B. Huang, A. Müller, J. Schnack, L.-S. Zheng, and Z. Zheng, *Chem. Commun.* **49**, 36 (2013)

- [73] A. Caneschi, A. Cornia, A.C. Fabretti, D. Gatteschi, and W. Malavasi, *Inorg. Chem.* **34**, 4660 (1995)
- [74] B. Pilawa, R. Desquiotz, M. Kelemen, M. Weickenmeier, and A. Geisselman, *J. Magn. Magn. Mater.* **177**, 748 (1997)
- [75] K. Bärwinkel, H.-J. Schmidt, and J. Schnack, *J. Magn. Magn. Mater.* **212**, 240 (2000)
- [76] W. Feller: *An introduction to probability theory and its applications*, Vol. 1 (John Wiley & Sons, New York, 3rd ed., 1968)
- [77] J. Schnack and R. Schnalle, *Polyhedron* **28**, 1620 (2009)
- [78] D. Kouzoudis, *J. Magn. Magn. Mater.* **173**, 259 (1997)
- [79] D. Kouzoudis, *J. Magn. Magn. Mater.* **189**, 366 (1998)
- [80] J. Bonner and M. Fisher, *Phys. Rev.* **135**, A640 (1964)
- [81] R. Botet and R. Jullien, *Phys. Rev. B* **27**, 631 (1983)
- [82] K. Fabricius, U. Löw, K.-H. Mütter, and P. Ueberholz, *Phys. Rev. B* **44**, 7476 (1991)
- [83] E. Manousakis, *Rev. Mod. Phys.* **63**, 1 (1991)
- [84] O. Golinelli, T. Jolicœur, and R. Lacaze, *Phys. Rev. B* **50**, 3037 (1994)
- [85] K. Fabricius, U. Löw, and J. Stolze, *Phys. Rev. B* **55**, 5833 (1997)
- [86] C. Lanczos, *J. Res. Nat. Bur. Stand.* **45**, 255 (1950)
- [87] J.K. Cullum and R.A. Willoughby:
Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I (Birkhäuser, Boston, 1985)
- [88] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst (Eds.):
Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide (Society for Industrial & Applied Mathematics, Philadelphia, 2000)
- [89] S.R. White and D. Huse, *Phys. Rev. B* **48**, 3844 (1993)
- [90] T. Xiang, *Phys. Rev. B* **58**, 9142 (1998)
- [91] I. Peschel, X. Wang, M. Kaulke, and K. Hallberg (Eds.): *Density-Matrix Renormalization* (Springer, Berlin, 1999)
- [92] J. Ummethum, J. Schnack, and A. Laeuchli, *J. Magn. Magn. Mater.* **327**, 103 (2013)
- [93] E. M. Stoudenmire and S. R. White, *Ann. Rev. Cond. Mat. Phys.* **3**, 111 (2012)

- [94] S. Liang and H. Pang, Phys. Rev. B **49**, 9214 (1994)
- [95] G.K.L. Chan and M. Head-Gordon, J. Chem. Phys. **116**, 4462 (2002)
- [96] E. Cuthill and J. McKee: in *Proceedings of the 1969 24th national conference* (ACM, New York, USA, 1969), pp. 157–172
- [97] W.M. Chan and A. George, BIT Numerical Mathematics **20**, 8 (1980)
- [98] S.W. Sloan: International Journal for Numerical Methods in Engineering **28**, 2651 (1989)
- [99] A. Albuquerque, F. Alet, P. Corboz, P. Dayal, A. Feiguin, S. Fuchs, L. Gamper, E. Gull, S. Gürtler, A. Honecker, R. Igarashi, M. Körner, A. Kozhevnikov, A. Läuchli, S. Manmana, M. Matsumoto, I. McCulloch, F. Michel, R. Noack, G. Pawłowski, L. Pollet, T. Pruschke, U. Schollwöck, S. Todo, S. Trebst, M. Troyer, P. Werner, and S. Wessel, J. Magn. Magn. Mater. **310**, 1187 (2007)
- [100] I. Rousochatzakis, A.M. Läuchli, and F. Mila, Phys. Rev. B **77**, 094420 (2008)
- [101] J. Schnack, M. Luban, and R. Modler, Europhys. Lett. **56**, 863 (2001)
- [102] J. Schnack, H.-J. Schmidt, J. Richter, and J. Schulenburg, Eur. Phys. J. B **24**, 475 (2001)
- [103] E. Neuscamman and G. K.-L. Chan, Phys. Rev. B **86**, 064402 (2012)
- [104] M. Aichhorn, M. Daghofer, H.G. Evertz, and W. von der Linden, Phys. Rev. B **67**, 161103 (2003)
- [105] J. Schnack, P. Hage, and H.-J. Schmidt, J. Comput. Phys. **227**, 4512 (2008)
- [106] N. Hoshino, M. Nakano, H. Nojiri, W. Wernsdorfer, and H. Oshio, J. Am. Chem. Soc. **131**, 15100 (2009)
- [107] W. Marshall, Proc. Royal. Soc. A (London) **232**, 48 (1955)
- [108] E.H. Lieb, T. Schultz, and D.C. Mattis, Ann. Phys. (N.Y.) **16**, 407 (1961)
- [109] E.H. Lieb and D.C. Mattis, J. Math. Phys. **3**, 749 (1962)
- [110] M. Karbach: *Finite-Size-Effekte im eindimensionalen Spin- $\frac{1}{2}$ -XXZ-Modell* (Ph.D. Thesis, Bergische Universität - Gesamthochschule Wuppertal, 1994)
- [111] K. Bärwinkel, H.-J. Schmidt, and J. Schnack, J. Magn. Magn. Mater. **220**, 227 (2000)
- [112] J. Schnack, Phys. Rev. B **62**, 14855 (2000)
- [113] K. Bärwinkel, P. Hage, H.-J. Schmidt, and J. Schnack, Phys. Rev. B **68**, 054422 (2003)
- [114] F. Haldane, Phys. Lett. A **93**, 464 (1983)

- [115] F. Haldane, Phys. Rev. Lett. **50**, 1153 (1983)
- [116] J. Schnack and M. Luban, Phys. Rev. B **63**, 014418 (2000)
- [117] O. Waldmann, Phys. Rev. B **65**, 024424 (2001)
- [118] B. Bernu, P. Lecheminant, C. Lhuillier, and L. Pierre, Phys. Rev. B **50**, 10048 (1994)
- [119] M. Gross, E. Sánchez-Velasco, and E.D. Siggia, Phys. Rev. B **40**, 11328 (1989)
- [120] L. Engelhardt and M. Luban, Phys. Rev. B **73**, 054430 (2006)
- [121] J. Richter, Phys. Rev. B **47**, 5794 (1993)
- [122] H.-J. Schmidt and M. Luban, J. Phys. A: Math. Gen. **36**, 6351 (2003)
- [123] J. Schulenburg, A. Honecker, J. Schnack, J. Richter, and H.-J. Schmidt, Phys. Rev. Lett. **88**, 167207 (2002)
- [124] J. Schnack, Dalton Trans. **39**, 4677 (2010)
- [125] H.-J. Schmidt, J. Phys. A: Math. Gen. **35**, 6545 (2002)
- [126] H.-J. Schmidt, J. Richter, and R. Moessner, J. Phys. A: Math. Gen. **39**, 10673 (2006)
- [127] H. Tasaki, Phys. Rev. Lett. **69**, 1608 (1992)
- [128] A. Mielke, J. Phys. A: Math. Gen. **25**, 4335 (1992)
- [129] A. Mielke and H. Tasaki, Commun. Math. Phys. **158**, 341 (1993)
- [130] O. Derzhko, A. Honecker, and J. Richter, Phys. Rev. B **76**, 220402 (2007)
- [131] O. Derzhko, A. Honecker, and J. Richter, Phys. Rev. B **79**, 054403 (2009)
- [132] W.F. Giaque and D. MacDougall, Phys. Rev. **43**, 768 (1933)
- [133] M.E. Zhitomirsky, Phys. Rev. B **67**, 104421 (2003)
- [134] L.J. Zhu, M. Garst, A. Rosch, and Q.M. Si, Phys. Rev. Lett. **91**, 066404 (2003)
- [135] M.E. Zhitomirsky and A. Honecker, J. Stat. Mech.: Theor. Exp. p. P07012 (2004)
- [136] O. Derzhko and J. Richter, Phys. Rev. B **70**, 104415 (2004)
- [137] J. Richter, J. Schulenburg, A. Honecker, J. Schnack, and H.-J. Schmidt, J. Phys.: Condens. Matter **16**, S779 (2004)
- [138] M.-J. Martinez-Perez, O. Montero, M. Evangelisti, F. Luis, J. Sese, S. Cardona-Serra, and E. Coronado, Adv. Mater. **24**, 4301 (2012)

9 Recent Advances in Experimental Research on High-Temperature Superconductivity

Bernhard Keimer

Max Planck Institute for Solid State Research

Heisenbergstr. 1, 70569 Stuttgart, Germany

Contents

1	Introduction	2
2	Magnetic order	4
3	Charge order	6
4	Spin fluctuations	8
5	Spin-fluctuation-mediated superconductivity	11
6	Challenges for theory	13
7	Outlook	14

1 Introduction

The theoretical description of electronic materials in the crossover regime between fully localized and fully itinerant electrons continues to be one of the greatest challenges in theoretical physics. The most profound problem in this field – the origin of high-temperature superconductivity – remains unsolved despite more than a quarter century of research. Recently, however, the experimental investigation of correlated-electron materials has made astounding progress, based on advances in materials synthesis and experimental methodology. As a result, the overall situation in this research field now looks very different than it did just ten years ago. Whereas some theoretical challenges appear less formidable than they did at that time, unforeseen new issues have been raised by the latest experiments. In this chapter, we will briefly summarize some of these developments, and then discuss some concrete challenges for theoretical research on cuprate superconductors in more detail. The emphasis will be on results obtained by spectroscopic methods.

A particularly influential development on the materials front has been the discovery and subsequent exploration of high-temperature superconductivity in iron pnictides and chalcogenides. Although these compounds exhibit a completely different chemical composition and lattice structure from the copper oxides, the phase diagrams of both classes of materials are closely related [1]. In particular, antiferromagnetically ordered phases at commensurate valence electron configuration are surrounded by superconducting phases at both lower and higher electron density. The observation of closely analogous low-energy spin fluctuations (including the so-called “resonant mode”) in the superconducting regimes of the phase diagram [2] has further highlighted the analogy to the cuprates and the case for magnetic mechanisms of Cooper pairing. At the same time, the antiferromagnetic state in the iron-based materials is a metallic spin density wave, rather than a Mott insulator with fully localized electrons. This implies that high-temperature superconductivity is not confined to “doped Mott insulators” – a class of systems in which the combination of strong correlations and disorder poses particularly profound theoretical problems. A theoretical approach to the high- T_c problem from the metallic limit with more effectively screened Coulomb interactions and a well-defined Fermi surface now appears much more promising than it did before the discovery of the iron-based superconductors.

A related breakthrough on the experimental front was made possible by the recently developed capability of carrying out transport and thermodynamic measurements in magnetic fields up to 100 T. Combined with the availability of single-crystal samples with very long transport mean free paths (including especially the stoichiometric underdoped compounds $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ and $\text{YBa}_2\text{Cu}_4\text{O}_8$), such experiments have led to the discovery of quantum oscillations indicative of Landau quasiparticles in underdoped cuprates [3–8], adding to earlier data on overdoped compounds [9, 10] above their respective critical fields. Quantum oscillations have also been observed in several iron pnictide superconductors, including the stoichiometric compounds LiFeAs and LiFeP [11]. Fermionic quasiparticles are thus a generic feature of high-temperature superconductors over a wide range of doping levels. This speaks in favor of theories based on Cooper pairing of conventional quasiparticles, and against various more exotic models of

high- T_c superconductivity.

The last decade also saw increasingly insightful experiments with scanning tunneling spectroscopy on correlated-electron materials, including the copper-based, iron-based, and heavy-fermion superconductors. These experiments opened our eyes to nanoscale inhomogeneities of the electron density induced by randomly placed dopant atoms in some of the most extensively investigated materials including superconducting $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$ and $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ [12,13]. Because of the low carrier density, the Coulomb potential of the ionized donor and acceptor atoms is much more poorly screened than in conventional metals. This can lead to a pronounced inhomogeneous broadening of spectroscopic features in volume-averaging experiments (including especially photoemission spectroscopy) on non-stoichiometric materials. These extrinsic effects must be considered before interpreting broad spectroscopic features as evidence of non-Fermi-liquid behavior. Disorder and inhomogeneity have also led to the development of a new technique, quasiparticle interference (QPI) spectroscopy [14], as a powerful phase-sensitive probe of the superconducting order parameter in both copper- [15] and iron-based [16] superconductors.

Further, the research field has benefitted greatly from the increase in energy resolution of spectroscopic probes such as inelastic neutron scattering (INS), angle-resolved photoemission spectroscopy (ARPES), and (non-resonant) inelastic x-ray scattering (IXS), which can now be performed with resolutions of $1\text{ }\mu\text{eV}$ in the former and 1 meV in the latter two. While high-resolution INS is yielding new insights into “spin freezing” phenomena in both copper oxides [17] and iron arsenides [18], high-resolution ARPES has not only led to a detailed description of the energy and momentum dependence of the superconducting gap function in high-temperature superconductors [19], but has recently also provided equivalent data sets for NbSe_2 and other classical charge density wave materials with smaller gaps [20,21]. Notably, these experiments have led to the discovery of a “pseudogap” and “Fermi arcs” above the CDW transition temperature in TaSe_2 and NbSe_2 [20,21]. High-resolution IXS experiments provide detailed insights into the role of the electron-phonon interaction in driving charge-density-wave formation [22–24]. Following these developments, spectroscopic data on cuprates [22] can now be calibrated against the behavior of their more conventional cousins [23,24].

Over the past five years, resonant elastic (REXS) and inelastic (RIXS) x-ray scattering have had a tremendous impact in research on correlated-electron systems. REXS allows the determination of spin, charge, and orbital order of the valence electron system with very high sensitivity [25], and has been instrumental for the recent discovery of charge density waves in bulk copper-oxide superconductors [26–30]. RIXS, on the other hand, was known ten years ago mainly as a momentum-resolved probe of interband transitions in the 1–5 eV range, quite similar to electron energy loss spectroscopy. Following a phenomenal improvement of the energy resolution by about an order of magnitude [31], RIXS experiments have resolved orbital and spin excitations in a variety of metal oxides, including recently the copper-oxide [32–35] and iron-pnictide [36] superconductors. With its high sensitivity to high-energy excitations, which is complementary to INS, RIXS is becoming an increasingly powerful spectroscopic probe of correlated-electron materials.

These developments on the experimental front have opened up various new theoretical challenges that could hardly be foreseen a decade ago. First, the theoretical foundation for some of the new techniques is far from complete, and it will sometimes go hand-in-hand with the understanding of the systems to be investigated. Whereas the expression of the non-resonant x-ray and neutron scattering cross sections in terms of density-density and spin-spin correlations is generally understood and accepted, an analogous formalism for REXS and RIXS is still under development. It is already clear that a comprehensive description of the photon energy dependence of the REXS and RIXS cross sections will have to take Coulomb interactions into account – the same interactions whose low-energy manifestation are the subject of investigation. Similarly, a comprehensive understanding of QPI in tunneling spectroscopy will require detailed information about the defects that scatter the quasiparticles. Even more fundamental challenges are raised by the theoretical description of pump-probe techniques that take the correlated-electron system far out of thermal equilibrium.

Here we will highlight some theoretical challenges from recent experiments that can be spelled out independent of technical details of the experimental probes.

2 Magnetic order

The generic phase diagram of the copper oxides (Fig. 1) includes a Mott-insulating phase centered around the doping level $p = 0$ corresponding to a single hole per Cu site in the CuO_2 planes, and a d -wave superconducting phase extending from $p \sim 0.05$ to ~ 0.25 . In the Mott-insulating phase, commensurate, collinear antiferromagnetic order is observed with ordering wave vector $\mathbf{q} = (\pi, \pi)$ (in a notation in which the nearest-neighbor lattice parameter $a \sim 3.8 \text{ \AA}$ is set to unity). Static magnetic order persists at low temperatures over some range of p , but the ordering wave vector is shifted away from (π, π) . In both $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ and $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, the two systems where the doping-induced commensurate-incommensurate transition has been studied in detail, the amplitude of the magnetization in this state remains a substantial fraction of the antiferromagnetic sublattice magnetization at $p = 0$, and the incommensurability δ increases monotonically with p [17, 39, 40]. The direction of the propagation vector in both compounds is different (along the Cu-O bond in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, and 45° away from it in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ for $0.02 \leq p \leq 0.05$).

High-resolution neutron scattering and muon spin rotation (μSR) studies of $\text{YBa}_2\text{Cu}_3\text{O}_{6.35}$ with $p \sim 0.06$ have demonstrated that the incommensurate magnetic order at wave vector $(\pi \pm \delta, \pi)$ is highly unstable to thermal fluctuations [17]. Even at temperatures of a few Kelvin, the signatures of static magnetic order in both sets of experiments are replaced by those of a slowly relaxing local magnetization. The “wipeout” of the nuclear magnetic resonance signals in other underdoped cuprates at low temperatures has also been attributed to slow spin fluctuations of this kind [41, 42]. This behavior is consistent with the “critical slowing down” expected in proximity to a zero-temperature phase transition in two-dimensional Heisenberg systems [43] in combination with disorder due to dopant atoms that limit the exponential divergence of the spin-spin correlation length at low temperatures [44].

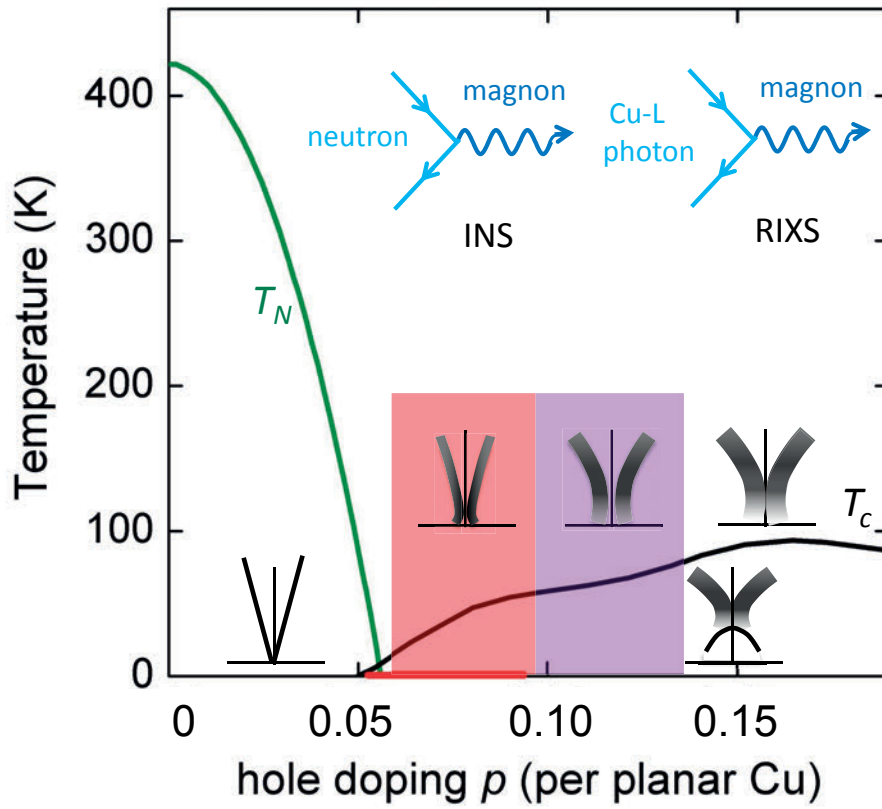


Fig. 1: Phase diagram of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. The Néel temperature, T_N , and superconducting transition temperature, T_c , were taken from Refs. [37] and [38], respectively. The red line indicates the stability range of static incommensurate magnetic order. Pink and purple shaded regions indicate temperature and doping regimes with low-energy incommensurate spin and charge correlations, respectively. The insets show diagrams illustrating INS and RIXS from spin excitations, as well as sketches of the dispersion and spectral weight distribution of the spin excitations around the antiferromagnetic ordering wavevector $\mathbf{q} = (\pi, \pi)$ in the different regimes of the phase diagram.

Upon further heating, low-energy spin fluctuations persist, but the incommensurability δ decreases continuously with increasing temperature. At temperatures exceeding $T \sim 150$ K, the signature of the uniaxial magnetic modulation is no longer visible in the neutron scattering data, and the magnetic response is centered at $\mathbf{q} = (\pi, \pi)$. The order-parameter-like temperature dependence of δ is consistent with a proximate “nematic” phase transition where the fourfold rotational symmetry of the CuO_2 layers is spontaneously broken [45]. In the orthorhombic crystal structure generated by the oxygen dopant atoms in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ with $x \geq 0.2$, this transition is expected to be broadened into a crossover. Thermodynamic singularities akin to those associated with a nematic transition in tetragonal $\text{Sr}_3\text{Ru}_2\text{O}_7$ [46] are indeed not observed in the cuprates. However, the observation of a similar temperature-driven incommensurate-commensurate transition in lightly doped $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ [47] shows that it reflects an intrinsic trend of the correlated electrons in the CuO_2 planes, rather than subtleties of the crystal structure of specific compounds such as the chains of oxygen dopant atoms in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. Uniaxially

modulated spin structures with unusual thermal properties thus appear to be a generic feature of all cuprates at doping levels close to the Mott-insulating phase.

In some compounds with intrinsically low superconducting T_c , including especially materials with composition $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ and $\text{La}_{2-x-y}\text{Sr}_x(\text{Nd,Eu})_y\text{CuO}_4$ (“214 compounds”) that exhibit the “low-temperature tetragonal” (LTT) crystal structure, uniaxial incommensurate magnetic order with wave vectors $\mathbf{q} = (\pi \pm \delta, \pi)$ and $(\pi, \pi \pm \delta)$ has also been observed at higher doping levels, in some cases up to $p \sim 0.15$ [48–55]. Corresponding charge-modulation peaks at $\mathbf{q} = (2\delta, 0)$ and $\mathbf{q} = (0, 2\delta)$ indicate that these magnetic peaks arise from a uniaxial (“stripe”) modulation of the spin amplitude. Static stripe order is only observed in compounds with the LTT structure whose primitive vectors are parallel to the stripe propagation vector – a situation that favors pinning to the lattice. Orthorhombic $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ does not exhibit stripe order, but low-energy spin fluctuations with the same momentum-space signature have been interpreted as evidence of fluctuating stripes [49, 51].

In all other cuprates investigated so far, static magnetic order disappears for $p \gtrsim 0.07$, and the magnetic excitation spectrum determined by neutron scattering develops a sizable spin gap [56], as expected for a magnetic quantum critical point. However, spinless Zn impurities in the CuO_2 planes nucleate static incommensurate magnetic order with a correlation length of a few unit cells in the spin-gap regime [57, 58]. Even at an impurity concentration of less than 1%, this leads to inhomogeneous coexistence between different electronic phases. Dopant-induced disorder, which is particularly pronounced in the 214 compounds [59], may contribute to the stability of the “stripe” phase in nominally pristine members of this family.

3 Charge order

Resonant [26, 28, 30] and nonresonant [27, 29] x-ray diffraction experiments on $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ samples in the doping regime $0.07 \leq p \leq 0.13$ have revealed biaxial, incommensurate charge density wave (CDW) correlations. Similar CDW correlations were demonstrated very recently in REXS experiments [60] on $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$, $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$, and $\text{HgBa}_2\text{CuO}_{4+\delta}$, indicating that incommensurate biaxial CDW correlations are a universal feature of the underdoped cuprates. The recent REXS experiments are qualitatively consistent with prior STS measurements that have revealed charge modulations in several cuprate families [61–63]. However, since the CDW features are superposed by electronic reconstructions induced by incommensurate lattice modulations in the former two compounds (some of which have turned out to be surface sensitive [64]), they are more difficult to interpret than those in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$.

The compilation of REXS data on $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ in Fig. 2 shows a continuous increase of the amplitude and a reduction of the wave vector of these correlations with increasing p . On a qualitative level, the doping dependence of the CDW wave vector (Fig. 3) tracks the distance between the antinodal regions of the Fermi surface (Fig. 4), which shrinks as the Fermi surface expands with increasing doping level. A quantitative comparison with ARPES data, however, reveals a better agreement with the distance between the tips of the “Fermi arcs” where the density of states is enhanced due to the opening of the pseudogap [60]. This suggests an intimate rela-

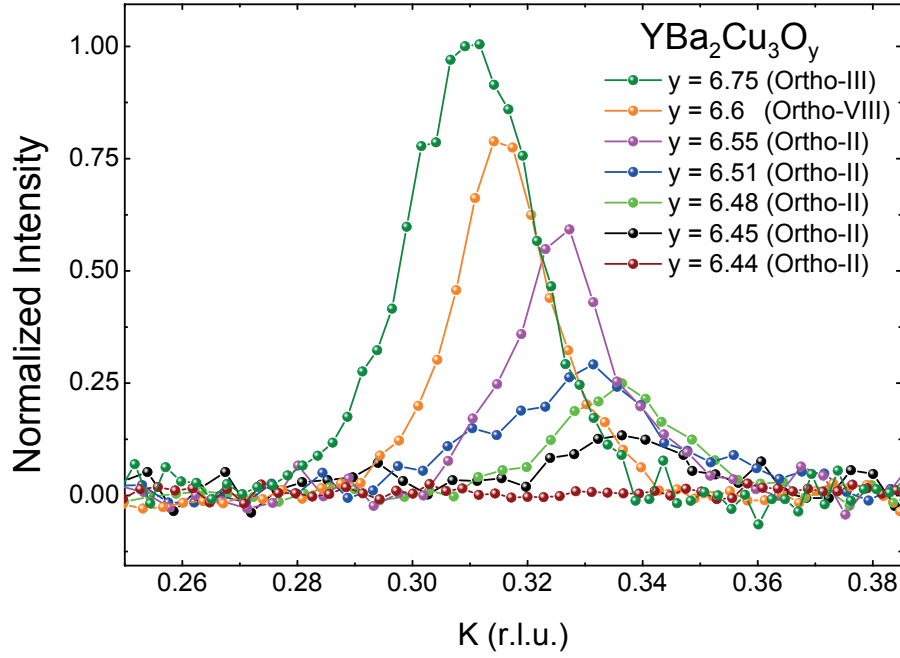


Fig. 2: REXS scans with photon energy tuned to the L -absorption edges of planar copper atoms in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. The scans are along $\mathbf{q} = (0, K)$ in the CuO_2 layers, where K is measured in reciprocal lattice units (r.l.u.). The labels Ortho-II, III, and VIII refer to the arrangement of oxygen dopant atoms [26, 28, 30].

relationship between the “Fermi arc” phenomenon and the CDW that should be further explored. Both the intensity and the correlation length of the CDW peaks grow upon cooling and exhibit pronounced maxima at the superconducting T_c [26–30]. This directly demonstrates competition between superconductivity and CDW formation, which predominantly affects electronic states near the antinodal regions of the Fermi surface. The strong competition between both types of order implied by these findings also explains the well-known plateau in the T_c -versus- p relation. Very recent high-resolution IXS data [22] have shown large anomalies of acoustic phonons associated with CDW formation, as well as a “central peak” indicative of static CDW regions nucleated by lattice defects. Pinning of soft phonons associated with structural phase transitions has also been observed in classical materials such as SrTiO_3 and Nb_3Sn , albeit over a much narrower temperature range. The persistence of this domain state over a much wider temperature range than corresponding phenomena in classical materials [65, 66] probably reflects the strong competition between CDW correlations and superconductivity. The nanoscale CDW domains revealed by these experiments will certainly contribute to the anomalous transport and thermodynamic properties of the underdoped cuprates. Both the central peak and the acoustic-phonon anomalies abruptly disappear in optimally doped $\text{YBa}_2\text{Cu}_3\text{O}_7$ [22], indicating that the nanoscale inhomogeneity is detrimental to high-temperature superconductivity. We note, however, that superconductivity-induced anomalies of the Cu-O bond-bending phonon around the CDW wave vector have also been observed in $\text{YBa}_2\text{Cu}_3\text{O}_7$ [67]. These findings indicate an underlying zero-temperature CDW critical point [68], which deserves further experimental and theoretical investigation.

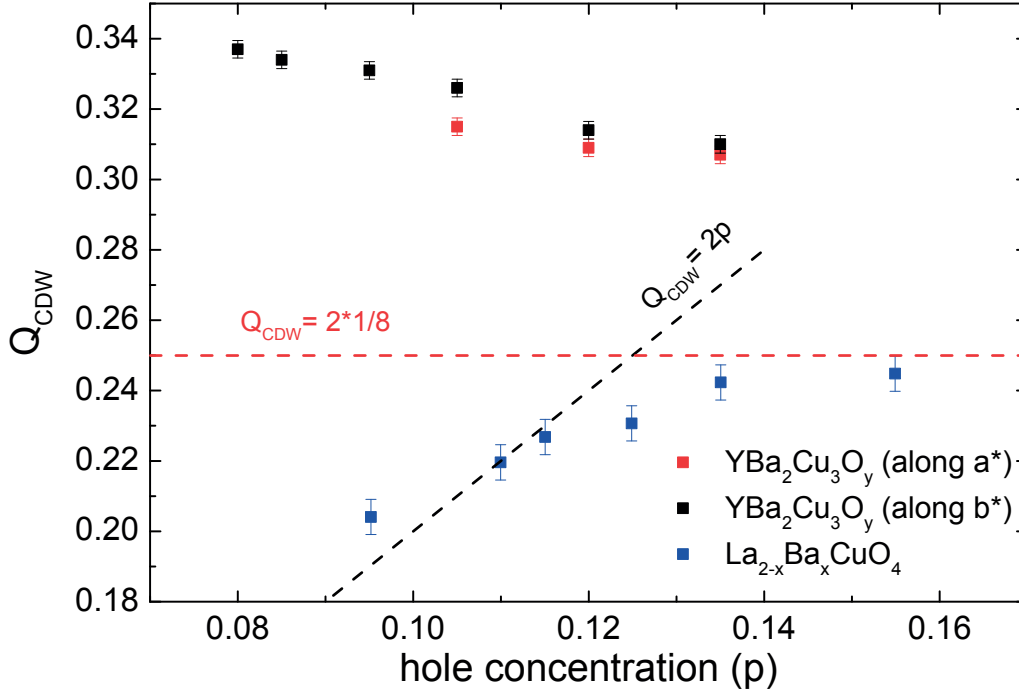


Fig. 3: Doping dependence of the CDW wavevector in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ [26, 28, 30], compared to the wave vector characterizing charge order in the “striped” state of $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ [55].

Whereas in zero magnetic field the competition between superconductivity and CDW order appears to preclude true CDW long-range order, recent NMR [69] and ultrasound [70] experiments have provided evidence of static CDW order and a thermodynamic CDW phase transition in high magnetic fields, where superconductivity is either greatly weakened or entirely obliterated. This is consistent with the reduction of the CDW peak intensity observed in REXS experiments in moderate magnetic fields [27, 29, 30]. In the presence of CDW long-range order, a Fermi surface reconstruction leading to the formation of small pockets is expected. Although this is an appealing explanation of the recent quantum oscillation data on stoichiometric $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ and $\text{YBa}_2\text{Cu}_4\text{O}_8$ [3, 7], a comprehensive, quantitatively consistent explanation of REXS and quantum oscillation data has not yet been reported.

4 Spin fluctuations

Since the discovery of the d -wave symmetry of the superconducting gap function, spin-fluctuation-mediated Cooper pairing has been one of the leading contenders in the quest for the mechanism of high-temperature superconductivity [1]. By combining INS and RIXS data, we now have detailed and comprehensive information about the spin fluctuation spectrum, which can be used as a basis for stringent tests of these models. We provide a brief survey of recent results on the doping evolution of the spin excitations, before discussing their implications for spin-fluctuation-mediated pairing theories.

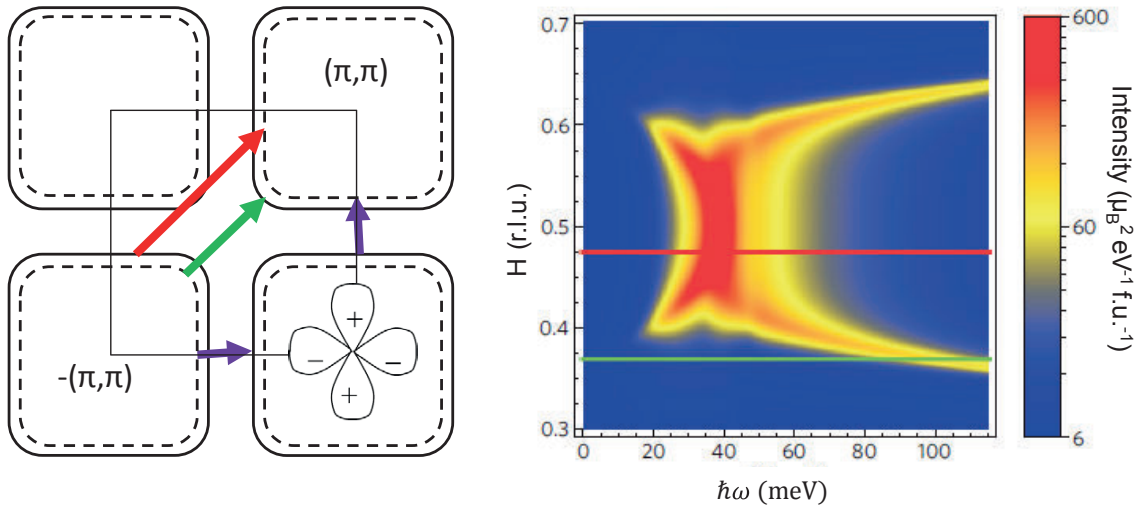


Fig. 4: Left panel: Kinematics of spin (red and green arrows) and charge (purple arrows) fluctuation scattering of quasiparticles on the Fermi surface of a bilayer compound such as $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$. Black solid and dashed lines correspond to Fermi surfaces in antibonding and bonding bands, respectively. The inset shows a sketch of the d -wave superconducting gap function. Right panel: Spin fluctuation intensity in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ along the (H, H) direction in the CuO_2 planes. H is measured in reciprocal lattice units, so that $H = 0.5$ corresponds to $\mathbf{q} = (\pi, \pi)$ [121].

In the Mott insulating state, the spin excitations determined by INS are well described as magnon modes of the 2D Heisenberg model with nearest-neighbor superexchange interaction $J \sim 130 - 140$ meV. The magnons emerge from the antiferromagnetic ordering wave vector $\mathbf{q} = (\pi, \pi)$ and are nearly gapless due to the weak magneto-crystalline anisotropy of the Cu ions (Fig. 1) [39,40]. In bilayer compounds such as $\text{YBa}_2\text{Cu}_3\text{O}_6$, an additional superexchange interaction between spins in directly adjacent CuO_2 layers, $J_\perp \sim 0.1J$, has to be taken into account. This leads to a non-generic optical magnon branch with a gap of 70 meV at $\mathbf{q} = (\pi, \pi)$ [71,72]. All other exchange interactions are significantly weaker.

In compounds with incommensurate magnetic order, including lightly doped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ and $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, as well as moderately doped, stripe-ordered $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ with $x = 1/8$, gapless spin excitations emerge from the magnetic Bragg reflections (Fig. 1). Contrary to the magnon excitations in the commensurate antiferromagnet, however, these excitations do not follow the predictions of the linear spin wave theory, which for incommensurate magnets predicts spin wave “cones” with approximately uniform spectral weights along their rims, at least at low energies [76]. Instead, intense low-energy excitations emanating from the incommensurate magnetic reflections first disperse towards $\mathbf{q} = (\pi, \pi)$, then bend over and approach the magnon branch of the insulating cuprates. With increasing excitation energy, cuts of the magnetic dispersion surface along the ordering wavevector thus resemble an “hourglass” with an open neck around $\mathbf{q} = (\pi, \pi)$ at $\hbar\omega \sim 30 - 50$ meV [73,75]. Similar hourglass dispersions have recently also been observed in insulating manganates [83] and cobaltates [84] with incommensurate magnetic order.

Whereas in the cuprates with 214 structure the magnetic excitations in the normal state remain nearly gapless even when static magnetic order is not present, the spectral weight of low-energy spin excitations in $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ and other compounds with higher maximal T_c is strongly reduced, so that the excitation spectrum can be approximated as a gapped version of the “open hourglass” spectrum in the state with incommensurate magnetic order (Fig. 1) [56, 74]. This is consistent with a quantum phase transition between magnetically ordered and quantum-disordered phases.

The momentum distribution of the spin excitation intensity is similar in underdoped cuprates with and without magnetic order. Below the neck of the hourglass, it exhibits a uniaxial anisotropy, reflecting the uniaxial nature of the (real or proximate) ground state [74]. At excitation energies above the neck, it displays fourfold symmetry. This aspect cannot be reproduced by calculations attributing the spin modulation to static “stripes” [77, 78], but it is correctly captured by models incorporating strong charge fluctuations on top of the “striped” background [79, 80], and by models of spin fluctuations in metals near a nematic instability [81]. Models based on a spiral ground state also reproduce many of the salient features of the magnetic excitations [82].

In the superconducting state of $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, the neck of the hourglass closes, and a sharp “resonant” mode with a downward dispersion is formed below the superconducting energy gap (Fig. 1). As a prominent signature of magnetically mediated Cooper pairing, this mode has been studied very extensively [2, 85–97]. Its energy increases with increasing p up to optimal doping, and decreases in the overdoped state, qualitatively following the “dome” in the T_c -versus- p relation. Similar observations have been made in other compounds with optimal T_c around 90 K, including $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ [98–100], $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ [101], and $\text{HgBa}_2\text{CuO}_{4+\delta}$ [102]. Magnetic resonant modes have also been observed in iron pnictide [106–108] and heavy-fermion superconductors [109–111], and thus appear generic to superconductors near an antiferromagnetic instability. In overdoped cuprates, the spectral weight of the low-energy spin excitations around $\mathbf{q} = (\pi, \pi)$, including the one of the resonant modes below T_c , is gradually reduced [103–105], and disappears entirely (to within the sensitivity of INS) in strongly overdoped $\text{La}_{1.78}\text{Sr}_{0.22}\text{CuO}_4$ ($p = 0.22$) [116, 117].

Based on these observations, and on related anomalies in fermionic spectral functions, the resonant mode has been attributed to a feedback effect of the Cooper pairing interaction on low-energy spin fluctuations that mediate the pairing interactions [1, 112, 121]. Note that the BCS coherence factors in the dynamical spin susceptibility extinguish the spectral weight of the mode, unless the sign of the superconducting gap function changes sign at the Fermi surface. The observation of the resonant mode has therefore been taken as evidence for d -wave pairing symmetry in the cuprates [1, 87, 113], and of s_{\pm} symmetry in the iron pnictides [106–108, 114, 115]. The spin dynamics for $\hbar\omega \lesssim 100$ meV thus reflects the ground state of the spin system, which strongly evolves with temperature and doping and is influenced by details of the crystal and electronic structure of different cuprate families. On the other hand, recent RIXS measurements on $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ over a wide range of doping levels have shown that spin fluctuations with excitation energies $\hbar\omega \gtrsim 100$ meV are weakly doping and temperature dependent, and that their

dispersion relations and energy-integrated spectral weights remain closely similar to antiferromagnetic magnons in Mott-insulating $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ [33]. Very recently, we have extended these experiments to highly overdoped $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ ($p = 0.27$), a compound that features a single, isolated CuO_2 plane per formula unit and very low intrinsic disorder [34]. In analogy to the spin excitations above the ordering temperature of insulating magnets, the “paramagnon” excitations in the cuprates are indicative of short-range correlations between localized spins. Since ARPES [118], angle-dependent magnetoresistance [9], and quantum oscillation [10] experiments on $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ have demonstrated canonical Fermi-liquid behavior with a Fermi surface that agrees quantitatively with the predictions of density functional theory, the persistence of magnon-like excitations well into the Fermi-liquid regime of the cuprates is highly surprising. Related results have been reported for highly overdoped $\text{La}_{2-x}\text{Ca}_x\text{CuO}_4$ [35].

5 Spin-fluctuation-mediated superconductivity

Are these spin excitations mediators of the Cooper pairing interaction, or are they simply bystanders? The appearance of the magnetic resonant mode in the superconducting state indicates that the low-energy excitations near $\mathbf{q} = (\pi, \pi)$ are intimately involved in the formation of the superconducting state. The absolute spectral weight of the resonant mode translates into an exchange energy that is comparable to the superconducting condensation energy [88,97,119]. The recently detected superconductivity-induced enhancement of the two-magnon Raman scattering cross section (which is dominated by high-energy spin excitations near the antiferromagnetic zone boundary) in $\text{HgBa}_2\text{CuO}_{4+\delta}$ indicates that higher-energy excitations also experience significant feedback effects [120]. Feedback effects over the entire paramagnon spectrum are not unexpected, because (unlike phonons in conventional superconductors) the spin excitations are generated by the same electrons that form the Cooper pairs. Converse evidence gleaned from photoemission [123] and optical [124–126] spectroscopies also indicate substantial coupling between conduction electrons and bosonic excitations with a bandwidth of ~ 300 meV (comparable to that of the paramagnon spectrum), although based on these data alone it is hard to determine whether or not this coupling contributes positively to the d -wave Cooper pairing interaction.

These results have encouraged us and our collaborators to go a step further towards a quantitative description of spin-fluctuation mediated Cooper pairing. Specifically, we have taken the experimentally measured spin fluctuation spectra of $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ [121] and $\text{YBa}_2\text{Cu}_3\text{O}_7$ [33] as input for Eliashberg calculations of the superconducting gap, Δ , and T_c . These calculations are carried out in the framework of the “spin-fermion” model, which treats the spin excitations in a manner analogous to phonons in conventional superconductors. Given the failure of similar calculations to quantitatively describe the microscopic parameters of conventional superconductors, it is not unreasonable to be skeptical about their prospects for these highly correlated materials. One has to keep in mind, however, that most conventional superconductors have complex Fermi surfaces and many phonon branches, and that the momentum-dependent electron-phonon interaction is difficult to measure or compute accurately. Uncertainties in either of these quan-

tities translate into large errors in Δ and T_c . In the cuprates, only a single electronic band and a single excitation branch is relevant near the Fermi surface (apart from minor complications in materials with multilayer structures), and information about the spin-fermion coupling can be extracted from spectroscopic data, greatly reducing the associated uncertainties.

The spin-fermion model ultimately needs to be rigorously justified. On a qualitative level, however, it appears promising at least for optimally doped and overdoped materials, whose spin excitations are gapped and do not exhibit quantum-critical behavior, and whose Fermi surfaces are not strongly affected by the “pseudogap”. Herein lies the importance of the recent RIXS experiments, which have supplied detailed information about the spin excitations in the optimally and overdoped regimes of materials with intrinsically high T_c [33,34]. Lightly doped materials with incommensurate magnetic order and 214 materials close to a “stripe” instability, which have been extensively characterized by INS [39], clearly require a different theoretical treatment.

In the Eliashberg calculations for $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$, spin excitations extending from the spin gap at $\mathbf{q} = (\pi, \pi)$ up to $\hbar\omega \sim 200$ meV, which scatter electrons between states in lobes of the d -wave superconducting gap function with opposite sign, are pair forming [33,34]. The highest-energy excitations near the antiferromagnetic zone boundary are indifferent to d -wave pairing, while those near $\mathbf{q} = 0$, which scatter electrons within the same lobe of the gap function, are pair breaking. Since the latter excitations have much lower spectral weight than those near $\mathbf{q} = (\pi, \pi)$, this pair-breaking effect only leads to a minor reduction of T_c . Because of kinematical constraints, RIXS detects predominantly low- \mathbf{q} spin excitations that are only weakly involved in Cooper pairing. The recently reported weak doping-dependence of the RIXS cross section in the overdoped regime [34,35], where T_c depends strongly on p , therefore does not contradict spin fluctuation mediated Cooper pairing models. We note that INS data do in fact show a strong decrease of the spectral weight of spin fluctuations near $\mathbf{q} = (\pi, \pi)$ at high p , consistent with a reduction of the Cooper pairing strength [116,117].

The Eliashberg calculations predict a superconducting transition temperature $T_c \sim 170$ K and gap $\Delta \sim 60$ meV, about a factor of two larger than the experimental observations for $\text{YBa}_2\text{Cu}_3\text{O}_7$ [33,121]. Interestingly, infrared conductivity experiments have provided evidence of superconducting fluctuations in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ up to temperatures of order 180 K [127], suggesting that the outcome of the mean-field Eliashberg calculations is physically meaningful. In view of the experimental data presented in the previous section, competition from CDW order is likely to be a major factor in the suppression of T_c from its mean-field value. The calculations also account for the “kinks” in the electronic band dispersions in a quantitatively consistent manner [121,122]. These results, along with other efforts along the same lines [128,129] give rise to the hope that a systematic improvement of these calculations – combined with more complete and accurate experimental data – will finally lead to a quantitative understanding of high- T_c superconductivity.

6 Challenges for theory

In summary, advances in experimental research have yielded a detailed, consistent picture of the phase behavior and spin dynamics of the cuprates over a wide range of doping levels. Although our knowledge is still far from complete – further experimental work is required especially in the overdoped regime – the data now at hand are already an excellent basis for the assessment of spin-fluctuation-mediated pairing models. The recent experimental results present the following specific challenges for theoretical research.

First, based on the most recent experimental data, the interplay between spin and charge excitations, as well as the origin of the “pseudogap” and “Fermi arc” phenomena and their relationship to the phase behavior of underdoped cuprates can now be addressed in a well defined and quantitative manner. Recent analytical calculations have already yielded interesting insights in this regard [128, 130]. Numerical work on strong-correlation models, which has shown that a momentum dependent pseudogap can arise as a consequence of on-site Coulomb interactions, should now be able to compute the momentum dependent charge susceptibility, and compare the results with experimental data. Corresponding data on more weakly correlated metals such as NbSe₂ are available for comparison [20, 21].

Consideration of the competition between CDW order and superconductivity should lead to a systematic improvement of the Eliashberg calculations for spin-fluctuation-mediated superconductivity [33, 121]. Along the same lines, theoretical work is required to explore the suitability and limitations of the spin-fermion model underlying these calculations from microscopic Hamiltonians such as the Hubbard model [1]. Analytical or numerical calculations of the Hubbard model should explore whether such models can explain the experimentally observed persistence of magnon-like excitations up to at least $p \sim 0.3$ [34].

The latest set of IXS measurements on acoustic-phonon energies and linewidths [22] show that a complete understanding of the underdoped cuprates will require detailed knowledge of the electron-phonon interaction. Combined with ARPES data on low-energy anomalies in the electronic band dispersions, the IXS data will enable a new approach to the evaluation of the electron-phonon coupling strength. Although earlier INS measurements on Cu-O bond-bending and -stretching vibrations at energies 40-80 meV had yielded evidence of substantial electron-phonon interactions [131–135], it has been difficult to correlate these data quantitatively with ARPES, because multiple closely spaced phonon modes and spin fluctuations are present in this energy range. The data on acoustic phonons now allow a detailed evaluation of the coupling constants in the different channels (*s*- and *d*-wave Cooper pairing and CDW formation).

Collective excitations (such as phasons and amplitudons) associated with charge density wave order, their hybridization with phonons, and their influence on thermodynamic and transport properties should be explored theoretically. Specific predictions for the RIXS cross section of such excitations will also be useful. The influence of disorder on incommensurate CDW correlations and on the thermodynamic and transport properties should be addressed, in a manner analogous to recent work on disordered incommensurate magnets [44, 57]. Specifically, it should be interesting to explore how disorder influences the nature of the thermodynamic

singularities associated with CDW formation.

What is the origin of the differences in spin- and charge-ordering patterns in the different cuprate families? The “striped” phase with combined uniaxial spin and charge order in the 214 materials, which has drawn an enormous amount of attention over the past two decades, now appears to be atypical for the cuprates, whereas a spin-gapped state with biaxial incommensurate charge order appears to be generic. Note, however, that the total amplitude of the charge modulation in the striped and CDW states is closely similar [137]. In addition to differences in the Fermi surface geometry (especially the nesting conditions) in different compounds, theoretical work should address the role of soft phonons associated with low-temperature structural phase transitions, which are unique to the 214 family [136].

Finally, both experimental and theoretical research is required to finally settle the issue of the theoretically predicted loop-current order with $\mathbf{q} = 0$ [138]. Progress in experimental research is required to resolve the apparent contradiction between elastic neutron scattering experiments (based on which discovery claims have been made), and NMR and μ SR experiments (which have led to null results). Likewise, theoretical work has led to divergent claims about the stability of loop-current order, which need to be conclusively resolved.

7 Outlook

This brief review shows that we have come a long way in our understanding of the electronic properties of the cuprates, and of the mechanism of high- T_c superconductivity. Of course, bona-fide predictions will be the ultimate test of any theoretical understanding we believe to have gained. The recently acquired capability to synthesize transition-metal oxide heterostructures and superlattices with atomic-scale precision [139, 140] now offers many new perspectives in this regard. Promising proposals include the manipulation of charge transfer across metal-oxide interfaces [141, 142], and “re-engineering” of the cuprate Hamiltonian in nickelate perovskites by strain and spatial confinement [143]. The latter proposal has stimulated efforts to control the effective dimensionality [144], orbital degeneracy [145], and phase behavior [146] in nickelates. Superconductivity has not yet been observed, but it is still early days.

Dynamical control of the cuprates by light stimulation is another field that is just beginning to be explored. Among the early achievements in this newly emerging field are photoinduced superconductivity in a stripe-ordered 214 compound [147] and the enhancement of T_c in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ [148]. With further development, controlled pumping of single excitations may ultimately develop into a powerful new way of testing theories of correlated-electron materials.

Acknowledgments

The work described in this manuscript is partly the result of the author’s collaboration with many researchers listed in the references below. The author is especially grateful to Mathieu Le Tacon and Vladimir Hinkov for help in preparing the figures.

References

- [1] For a recent review, see D.J. Scalapino, *Rev. Mod. Phys.* **84**, 1383 (2012)
- [2] For a review, see Y. Sidis, S. Pailhès, B. Keimer, P. Bourges, C. Ulrich, and L.P. Regnault, *Phys. stat. sol. (b)* **241**, 1204 (2004)
- [3] For a recent review, S.E. Sebastian, N. Harrison, and G.G. Lonzarich, *Rep. Progr. Phys.* **75**, 102501 (2012)
- [4] N. Doiron-Leyraud, C. Proust, D. LeBoeuf, J. Levallois, J.-B. Bonnemaïson, R. Liang, D.A. Bonn, W.N. Hardy, L. Taillefer, *Nature* **447**, 566 (2007)
- [5] A.F. Bangura, J.D. Fletcher, A. Carrington, J. Levallois, M. Nardone, B. Vignolle, P.J. Heard, N. Doiron-Leyraud, D. LeBoeuf, L. Taillefer, S. Adachi, C. Proust, N.E. Hussey, *Phys. Rev. Lett.* **100**, 047004 (2008)
- [6] S.E. Sebastian, N. Harrison, R. Liang, D.A. Bonn, W.N. Hardy, C.H. Mielke, and G.G. Lonzarich, *Phys. Rev. Lett.* **108**, 196403 (2012)
- [7] F. Laliberté, J. Chang, N. Doiron-Leyraud, E. Hassinger, R. Daou, M. Rondeau, B.J. Ramshaw, R. Liang, D.A. Bonn, W.N. Hardy, S. Pyon, T. Takayama, H. Takagi, I. Sheikin, L. Malone, C. Proust, K. Behnia, L. Taillefer, *Nature Com.* **2**, 432 (2011)
- [8] S.C. Riggs, O. Vafek, J.B. Kemper, J.B. Betts, A. Migliori, F.F. Balakirev, W.N. Hardy, R. Liang, D.A. Bonn, G.S. Boebinger, *Nature Phys.* **7**, 332 (2011)
- [9] N.E. Hussey, M. Abdel-Jawad, A. Carrington, A.P. Mackenzie, and L. Balicas, *Nature* **425**, 814 (2003)
- [10] B. Vignolle, A. Carrington, R.A. Cooper, M.M. J. French, A.P. Mackenzie, C. Jaudet, D. Vignolles, C. Proust, and N.E. Hussey, *Nature* **455**, 952 (2008)
- [11] C. Putzke, A.I. Coldea, I. Guillaumon, D. Vignolles, A. McCollam, D. LeBoeuf, M.D. Watson, I.I. Mazin, S. Kasahara, T. Terashima, T. Shibauchi, Y. Matsuda, and A. Carrington, *Phys. Rev. Lett.* **108**, 047002 (2012)
- [12] S.H. Pan, J.P. O'Neal, R.L. Badzey, C. Chamon, H. Ding, J.R. Engelbrecht, Z. Wang, H. Eisaki, S. Uchida, A.K. Gupta, K.-W. Ng, E.W. Hudson, K.M. Lang, J.C. Davis, *Nature* **413**, 282 (2001)
- [13] M.C. Boyer, W.D. Wise, K. Chatterjee, Ming Yi, T. Kondo, T. Takeuchi, H. Ikuta, E.W. Hudson, *Nature Phys.* **3**, 802 (2007)
- [14] K. McElroy, R.W. Simmonds, J.E. Hoffman, D.-H. Lee, J. Orenstein, H. Eisaki, S. Uchida, J. C. Davis, *Nature* **422**, 592 (2003)

- [15] T. Hanaguri, Y. Kohsaka, M. Ono, M. Maltseva, P. Coleman, I. Yamada, M. Azuma, M. Takano, K. Ohishi, H. Takagi, *Science* **323**, 923 (2009)
- [16] T. Hanaguri, S. Niitaka, K. Kuroki, H. Takagi, *Science* **328**, 474 (2010)
- [17] D. Haug, V. Hinkov, Y. Sidis, P. Bourges, N.B. Christensen, A. Ivanov, T. Keller, C.T. Lin, B. Keimer, *New J. Phys.* **12**, 105006 (2010)
- [18] D.S. Inosov, G. Friemel, J.T. Park, A.C. Walters, Y. Texier, Y. Laplace, J. Bobroff, V. Hinkov, D.L. Sun, Y. Liu, R. Khasanov, K. Sedlak, Ph. Bourges, Y. Sidis, A. Ivanov, C.T. Lin, T. Keller, B. Keimer, *Phys. Rev. B* **87**, 224425 (2013)
- [19] For a recent review, see I.M. Vishik, W.S. Lee, R.-H. He, M. Hashimoto, Z. Hussain, T.P. Devereaux, and Z.X. Shen, *New J. Phys.* **12**, 105008 (2010)
- [20] S.V. Borisenko, A.A. Kordyuk, A.N. Yaresko, V.B. Zabolotnyy, D.S. Inosov, R. Schuster, B. Büchner, R. Weber, R. Follath, L. Patthey, and H. Berger, *Phys. Rev. Lett.* **100**, 196402 (2008)
- [21] S.V. Borisenko, A.A. Kordyuk, V.B. Zabolotnyy, D.S. Inosov, D. Evtushinsky, B. Büchner, A.N. Yaresko, A. Varykhalov, R. Follath, W. Eberhardt, L. Patthey, and H. Berger, *Phys. Rev. Lett.* **102**, 166402 (2009)
- [22] M. Le Tacon, A. Bosak, S.M. Souliou, G. Dellea, T. Loew, R. Heid, K.-P. Bohnen, G. Ghiringhelli, M. Krisch, B. Keimer, arXiv:1307.1673.
- [23] M. Hoesch, A. Bosak, D. Chernyshov, H. Berger, M. Krisch, *Phys. Rev. Lett.* **102**, 086402 (2009)
- [24] M. Leroux, M. Le Tacon, M. Calandra, L. Cario, M.-A. Méasson, P. Diener, E. Borrisenko, A. Bosak, and P. Rodière, *Phys. Rev. B* **86**, 155125 (2012)
- [25] For a review, see J. Fink, E. Schierle, E. Weschke, J. Geck, *Rep. Progr. Phys.* **76**, 056502 (2013)
- [26] G. Ghiringhelli, M. Le Tacon, M. Minola, S. Blanco-Canosa, C. Mazzoli, N.B. Brookes, G.M. De Luca, A. Frano, D.G. Hawthorn, F. He, T. Loew, M.M. Sala, D.C. Peets, M. Saluzzo, E. Schierle, R. Sutarto, G.A. Sawatzky, E. Weschke, B. Keimer, L. Braicovich, *Science* **337**, 821 (2012)
- [27] J. Chang, E. Blackburn, A.T. Holmes, N.B. Christensen, J. Larsen, J. Mesot, R. Liang, D.A. Bonn, W.N. Hardy, A. Watenphul, M. v. Zimmermann, E.M. Forgan, S.M. Hayden, *Nat. Phys.* **8**, 871 (2012)
- [28] A.J. Achkar, R. Sutarto, X. Mao, F. He, A. Frano, S. Blanco-Canosa, M. Le Tacon, G. Ghiringhelli, L. Braicovich, M. Minola, M. Moretti Sala, C. Mazzoli, R. Liang,

- D.A. Bonn, W.N. Hardy, B. Keimer, G.A. Sawatzky, D.G. Hawthorn,
Phys. Rev. Lett. **109**, 167001 (2012)
- [29] E. Blackburn, J. Chang, M. Hücker, A.T. Holmes, N.B. Christensen, R. Liang,
D.A. Bonn, W.N. Hardy, M. v. Zimmermann, E.M. Forgan, S.M. Hayden,
Phys. Rev. Lett. **110**, 137004 (2013)
- [30] S. Blanco-Canosa, A. Frano, T. Loew, Y. Lu, J. Porras, G. Ghiringhelli, M. Minola,
C. Mazzoli, L. Braicovich, E. Schierle, E. Weschke, M. Le Tacon, B. Keimer,
Phys. Rev. Lett. **110**, 187001 (2013)
- [31] L.J.P. Ament, M. van Veenendaal, T.P. Devereaux, J.P. Hill, J. van den Brink,
Rev. Mod. Phys. **83**, 705 (2011)
- [32] L. Braicovich, J. van den Brink, V. Bisogni, M.M. Sala, L.J.P. Ament, N.B. Brookes,
G.M. De Luca, M. Salluzzo, T. Schmitt, V.N. Strocov, G. Ghiringhelli,
Phys. Rev. Lett. **104**, 077002 (2010)
- [33] M. Le Tacon, G. Ghiringhelli, J. Chaloupka, M.M. Sala, V. Hinkov, M.W. Haverkort,
M. Minola, M. Bakr, K.J. Zhou, S. Blanco-Canosa, C. Monney, Y.T. Song, G.L.
Sun, C.T. Lin, G.M. De Luca, M. Salluzzo, G. Khaliullin, T. Schmitt, L. Braicovich,
B. Keimer, Nat. Phys. **7**, 725 (2011)
- [34] M. Le Tacon, M. Minola, D.C. Peets, M. Moretti Sala, S. Blanco-Canosa, V. Hinkov,
R. Liang, D.A. Bonn, W.N. Hardy, C.T. Lin, T. Schmitt, L. Braicovich, G. Ghiringhelli,
B. Keimer, Phys. Rev. B **88**, 020501(R) (2013)
- [35] M.P.M. Dean *et al.*, arXiv:1303.5359
- [36] K.-J. Zhou, Y.-B. Huang, C. Monney, X. Dai, V.N. Strocov, N.-L. Wang, Z.-G. Chen,
C. Zhang, P. Dai, L. Patthey, J. van den Brink, H. Ding, T. Schmitt,
Nat. Commun. **4**, 1470 (2013)
- [37] F. Coneri, S. Sanna, K. Zheng, J. Lord, R. De Renzi, Phys. Rev. B **81**, 104507 (2010)
- [38] R. Liang, D.A. Bonn, W.N. Hardy, Phys. Rev. B **73**, 180505 (2006)
- [39] For a review, see R.J. Birgeneau, C. Stock, J.M. Tranquada, K. Yamada,
J. Phys. Soc. Jpn. **75**, 111003 (2006)
- [40] For a review, see M. Fujita, H. Hiraka, M. Matsuda, M. Matsuura, J. M. Tranquada,
S. Wakimoto, G.Y. Xu, K. Yamada, J. Phys. Soc. Jpn. **81**, 011007 (2012)
- [41] M.-H. Julien, A. Campana, A. Rigamonti, P. Carretta, F. Borsa, P. Kuhns, A.P. Reyes,
W.G. Moulton, M. Horvatic, C. Berthier, A. Vietkin, A. Revcolevschi,
Phys. Rev. B **63**, 144508 (2001)

- [42] T. Wu, H. Mayaffre, S. Krämer, M. Horvatic, C. Berthier, C.T. Lin, D. Haug, T. Loew, V. Hinkov, B. Keimer, M.-H. Julien, *Phys. Rev. B* **88**, 014511 (2013)
- [43] E. Katanin, O.P. Sushkov, *Phys. Rev. B* **83**, 094426 (2011)
- [44] E.C. Andrade, M. Vojta, *EPL* **97**, 37007 (2011)
- [45] V. Hinkov, D. Haug, B. Fauqué, P. Bourges, Y. Sidis, A. Ivanov, C. Bernhard, C.T. Lin, B. Keimer, *Science* **319**, 597 (2008)
- [46] R.A. Borzi, S.A. Grigera, J. Farrell, R.S. Perry, S.J.S. Lister, S.L. Lee, D.A. Tennant, Y. Maeno, A.P. Mackenzie, *Science* **315**, 214 (2007)
- [47] M. Matsuda, J.A. Fernandez-Baca, M. Fujita, K. Yamada, J.M. Tranquada, *Phys. Rev. B* **84**, 104524 (2011)
- [48] J.M. Tranquada, B.J. Sternlieb, J.D. Axe, Y. Nakamura, S. Uchida, *Nature* **375**, 561 (1995)
- [49] For a review, see S.A. Kivelson, I.P. Bindloss, E. Fradkin, V. Oganessian, J.M. Tranquada, A. Kapitulnik, C. Howald, *Rev. Mod. Phys.* **75**, 1201 (2003)
- [50] P. Abbamonte, A. Rusydi, S. Smadici, G.D. Gu, G.A. Sawatzky, and D.L. Feng, *Nat. Phys.* **1**, 155 (2005)
- [51] For a review, see M. Vojta, *Adv. Phys.* **58**, 699 (2009)
- [52] J. Fink, E. Schierle, E. Weschke, J. Geck, D. Hawthorn, V. Soltwisch, H. Wadati, H.-H. Wu, H.A. Dürr, N. Wizent, B. Büchner, G.A. Sawatzky, *Phys. Rev. B* **79**, 100502 (2009)
- [53] J. Fink, V. Soltwisch, J. Geck, E. Schierle, E. Weschke, and B. Büchner, *Phys. Rev. B* **83**, 092503 (2011)
- [54] S.B. Wilkins, M.P.M. Dean, J. Fink, M. Hücker, J. Geck, V. Soltwisch, E. Schierle, E. Weschke, G. Gu, S. Uchida, N. Ichikawa, J.M. Tranquada, J.P. Hill, *Phys. Rev. B* **84**, 195101 (2011)
- [55] M. Hücker, M. v. Zimmermann, G.D. Gu, Z.J. Xu, J.S. Wen, G. Xu, H.J. Kang, A. Zheludev, J.M. Tranquada, *Phys. Rev. B* **83**, 104506 (2011)
- [56] H.F. Fong, P. Bourges, Y. Sidis, L.P. Regnault, J. Bossy, A. Ivanov, D.L. Milius, I.A. Aksay, B. Keimer, *Phys. Rev. B* **61**, 14773 (2000)
- [57] H. Alloul, J. Bobroff, M. Gabay, P.J. Hirschfeld, *Rev. Mod. Phys.* **81**, 45 (2009)
- [58] A. Suchaneck, V. Hinkov, D. Haug, L. Schulz, C. Bernhard, A. Ivanov, K. Hradil, C.T. Lin, P. Bourges, B. Keimer, and Y. Sidis, *Phys. Rev. Lett.* **105**, 037207 (2010)

- [59] W. Chen, G. Khaliullin, O.P. Sushkov, Phys. Rev. B **80**, 094519 (2009)
- [60] R. Comin *et al.*, submitted; E.H. da Silva Neto *et al.*, submitted; M. Greven *et al.*, unpublished data.
- [61] W.D. Wise, M.C. Boyer, K. Chatterjee, T. Kondo, T. Takeuchi, H. Ikuta, Y. Wang, E.W. Hudson, Nat. Phys. **4**, 696 (2008)
- [62] Y. Kohsaka, C. Taylor, K. Fujita, A. Schmidt, C. Lupien, T. Hanaguri, M. Azuma, M. Takano, H. Eisaki, H. Takagi, S. Uchida, and J. C. Davis, Science **315**, 1380 (2007)
- [63] C.V. Parker, P. Aynajian, E.H. da Silva Neto, A. Pushp, S. Ono, J. Wen, Z. Xu, G. Gu, and A. Yazdani, Nature **468**, 677 (2010)
- [64] J.A. Rosen, R. Comin, G. Levy, D. Fournier, Z.-H. Zhu, B. Ludbrook, C.N. Veenstra, A. Nicolaou, D. Wong, P. Dosanjh, Y. Yoshida, H. Eisaki, G.R. Blake, F. White, T.T.M. Palstra, R. Sutarto, F. He, A. Fraño Pereira, Y. Lu, B. Keimer, G. Sawatzky, L. Petaccia, A. Damascelli, Nature Com. **4**, 1977 (2013)
- [65] R.J. Cowley, Phil. Trans. R. Soc. Lond. A **354**, 2799 (1996)
- [66] J.D. Axe, G. Shirane, Phys. Rev. B **8**, 1965 (1973)
- [67] M. Raichle, D. Reznik, D. Lamago, R. Heid, Y. Li, M. Bakr, C. Ulrich, V. Hinkov, K. Hradil, C.T. Lin, B. Keimer, Phys. Rev. Lett. **107**, 177004 (2011)
- [68] C. Castellani, C. Di Castro, M. Grilli, Phys. Rev. Lett. **75**, 4650 (1995)
- [69] T. Wu, H. Mayaffre, S. Krämer, M. Horvatic, C. Berthier, W. N. Hardy, R. Liang, D.A. Bonn, M.-H. Julien, Nature **477**, 191 (2011)
- [70] D. LeBoeuf, S. Krämer, W.N. Hardy, R. Liang, D.A. Bonn, and C. Proust, Nat. Phys. **8**, 871 (2012)
- [71] D. Reznik, H.F. Fong, L.P. Regnault, J. Bossy, C. Vettier, D.L. Milius, I.A. Aksay, B. Keimer, Phys. Rev. B **53**, R14741 (1996)
- [72] S.M. Hayden, G. Aeppli, T.G. Perring, H.A. Mook, F. Dogan, Phys. Rev. B **54**, R6905 (1996)
- [73] J.M. Tranquada, H. Woo, T.G. Perring, H. Goka, G.D. Gu, G. Xu, M. Fujita, K. Yamada, Nature **429**, 534 (2004)
- [74] V. Hinkov, P. Bourges, S. Pailh  s, Y. Sidis, A. Ivanov, C.D. Frost, T.G. Perring, C.T. Lin, D.P. Chen B. Keimer, Nature Phys. **3**, 780 (2007)
- [75] O.J. Lipscombe, B. Vignolle, T.G. Perring, C.D. Frost, S.M. Hayden, Phys. Rev. Lett. **102**, 167002 (2009)

- [76] F. Krüger, S. Scheidl, Phys. Rev. B **67**, 134512 (2003)
- [77] G.S. Uhrig, K.P. Schmidt, M. Grüninger, Phys. Rev. Lett. **93**, 267003 (2004)
- [78] D.X. Yao, E.W. Carlson, D.K. Campbell, Phys. Rev. B **73**, 224525 (2006)
- [79] M. Vojta, T. Vojta, R.K. Kaul, Phys. Rev. Lett. **97**, 097001 (2006)
- [80] G. Seibold, J. Lorenzana, Phys. Rev. B **73**, 144515 (2006)
- [81] H. Yamase, W. Metzner, Phys. Rev. B **73**, 214517 (2006)
- [82] A.I. Milstein, O.P. Sushkov, Phys. Rev. B **78**, 014501 (2008)
- [83] H. Ulbrich, P. Steffens, D. Lamago, Y. Sidis, M. Braden, Phys. Rev. Lett. **108**, 247209 (2012)
- [84] A.T. Boothroyd, P. Babkevich, D. Prabhakaran, P.G. Freeman, Nature **471**, 341 (2011)
- [85] J. Rossat-Mignod, L.P. Regnault, C. Vettier, P. Bourges, P. Burlet, J. Bossy, J.Y. Henry, G. Lapertot, Physica C **185**, 86 (1991)
- [86] H.A. Mook, M. Yethiraj, G. Aeppli, T.E. Mason, T. Armstrong, Phys. Rev. Lett. **70**, 3490(1993)
- [87] H.F. Fong, B. Keimer, P.W. Anderson, D. Reznik, F. Dogan, I.A. Aksay, Phys. Rev. Lett. **75**, 316 (1995)
- [88] H.F. Fong, B. Keimer, D. Reznik, D.L. Milius, I.A. Aksay, Phys. Rev. B **54**, 6708 (1996)
- [89] P.C. Dai, H.A. Mook, S.M. Hayden, G. Aeppli, T.G. Perring, R.D. Hunt, F. Dogan, Science **284**, 1344 (1999)
- [90] P. Bourges, Y. Sidis, H. F. Fong, L.P. Regnault, J. Bossy, A. Ivanov, B. Keimer, Science **288**, 1234 (2000)
- [91] S. Pailhès, Y. Sidis, P. Bourges, V. Hinkov, A. Ivanov, C. Ulrich, L.P. Regnault, B. Keimer, Phys. Rev. Lett. **93**, 167001 (2004)
- [92] S.M. Hayden, H.A. Mook, P. Dai, T.G. Perring, F. Dogan, Nature **429**, 531 (2004).
- [93] V. Hinkov, S. Pailhès, P. Bourges, Y. Sidis, A. Ivanov, A. Kulakov, C. T. Lin, D. P. Chen, C. Bernhard, B. Keimer, Nature **430**, 650 (2004)
- [94] C. Stock, W.J.L. Buyers, R. Liang, D. Peets, Z. Tun, D. Bonn, W.N. Hardy, R.J. Birgeneau, Phys. Rev. B **69**, 014502 (2004)
- [95] C. Stock, W.J.L. Buyers, R.A. Cowley, P.S. Clegg, R. Coldea, C.D. Frost, R. Liang, D. Peets, D. Bonn, W.N. Hardy, R.J. Birgeneau, Phys. Rev. B **71**, 024522 (2005)

- [96] Pailhès, C. Ulrich, B. Fauqué, V. Hinkov, Y. Sidis, A. Ivanov, C.T. Lin, B. Keimer, P. Bourges, *Phys. Rev. Lett.* **96**, 257001 (2006)
- [97] H. Woo, P. Dai, S.M. Hayden, H.A. Mook, T. Dahm, D.J. Scalapino, T.G. Perring, F. Dogan, *Nature Phys.* **2**, 600 (2006)
- [98] H.F. Fong, P. Bourges, Y. Sidis, L.P. Regnault, A. Ivanov, G.D. Gu, N. Koshizuka, B. Keimer, *Nature* **398**, 588 (1999)
- [99] B. Fauqué, Y. Sidis, L. Capogna, A. Ivanov, K. Hradil, C. Ulrich, A.I. Rykov, B. Keimer, P. Bourges. *Phys. Rev. B* **76**, 214512 (2007)
- [100] G. Xu, G.D. Gu, M. Hücker, B. Fauqué, T.G. Perring, L.P. Regnault, J.M. Tranquada, *Nature Phys.* **5**, 642 (2009)
- [101] H. He, P. Bourges, Y. Sidis, C. Ulrich, L.P. Regnault, S. Pailhès, N.S. Berzigiarova, N.N. Kolesnikov, B. Keimer, *Science* **295**, 1045 (2002)
- [102] G. Yu, Y. Li, E. M. Motoyama, X. Zhao, N. Barisic, Y.Cho, P. Bourges, K. Hradil, R. A. Mole, M. Greven, *Phys. Rev. B* **81**, 064518 (2010)
- [103] H. He, Y. Sidis, P. Bourges, G.D. Gu, A. Ivanov, N. Koshizuka, B. Liang, C. T. Lin, L.P. Regnault, E. Schoenherr, B. Keimer, *Phys. Rev. Lett.* **86**, 1610 (2001)
- [104] S. Pailhès, Y. Sidis, P. Bourges, C. Ulrich, V. Hinkov, L.P. Regnault, A. Ivanov, B. Liang, C.T. Lin, C. Bernhard, B. Keimer, *Phys. Rev. Lett.* **91**, 237002 (2003)
- [105] L. Capogna, B. Fauqué, Y. Sidis, C. Ulrich, P. Bourges, S. Pailhès, A. Ivanov, J.L. Tallon, B. Liang, C.T. Lin, A.I. Rykov, B. Keimer, *Phys. Rev. B* **75**, 060502(R) (2007)
- [106] A.D. Christianson, E.A. Goremychkin, R. Osborn, S. Rosenkranz, M.D. Lumsden, C.D. Malliakas, I.S. Todorov, H. Claus, D.Y. Chung, M.G. Kanatzidis, R.I. Bewley, T. Guidi, *Nature* **456**, 930 (2008)
- [107] M.D. Lumsden, A.D. Christianson, D. Parshall, M.B. Stone, S.E. Nagler, G.J. MacDougall, H.A. Mook, K. Lokshin, T. Egami, D.L. Abernathy, E.A. Goremychkin, R. Osborn, M.A. McGuire, A.S. Sefat, R. Jin, B.C. Sales, D. Mandrus, *Phys. Rev. Lett.* **102**, 107005 (2009)
- [108] D.S. Inosov, J.T. Park, P. Bourges, D.L. Sun, Y. Sidis, A. Schneidewind, K. Hradil, D. Haug, C.T. Lin, B. Keimer, V. Hinkov, *Nature Phys.* **6**, 178 (2010)
- [109] N.K. Sato, N. Aso, K. Miyake, R. Shiina, P. Thalmeier, G. Varelogiannis, C. Geibel, F. Steglich, P. Fulde, and T. Komatsubara, *Nature* **410**, 340 (2001)
- [110] C. Stock, C. Broholm, J. Hudis, H.J. Kang, and C. Petrovic, *Phys. Rev. Lett.* **100**, 087001 (2008)

- [111] O. Stockert, J. Arndt, E. Faulhaber, C. Geibel, H.S. Jeevan, J. Kirchner, S.M. Loewenhaupt, K. Schmalzl, W. Schmidt, Q. Si, F. Steglich, *Nature Phys.* **7**, 119 (2011)
- [112] M. Eschrig, *Adv. Phys.* **55**, 47 (2006)
- [113] P. Monthoux, D.J. Scalapino, *Phys. Rev. Lett.* **72**, 1874 (1994)
- [114] M.M. Korshunov, I. Eremin, *Phys. Rev. B* **78**, 140509 (2008)
- [115] T.A. Maier, S. Graser, D.J. Scalapino, P. Hirschfeld, *Phys. Rev. B* **79**, 134520 (2009)
- [116] S. Wakimoto, K. Yamada, J. M. Tranquada, C.D. Frost, R.J. Birgeneau, H. Zhang, *Phys. Rev. Lett.* **98**, 247003 (2007)
- [117] O.J. Lipscombe, S.M. Hayden, B. Vignolle, D.F. McMorrow, T.G. Perring, *Phys. Rev. Lett.* **99**, 067002 (2007)
- [118] M. Platié, J.D.F. Mottershead, I.S. Elfimov, D.C. Peets, R. Liang, D.A. Bonn, W.N. Hardy, S. Chiuzbaian, M. Falub, M. Shi, L. Patthey, A. Damascelli, *Phys. Rev. Lett.* **95**, 077001 (2005)
- [119] E. Demler, S.C. Zhang, *Nature* **396**, 733 (1998)
- [120] Y. Li, M. Le Tacon, M. Bakr, D. Terrade, D. Manske, R. Hackl, L. Ji, M. K. Chan, N. Barisic, X. Zhao, M. Greven, B. Keimer, *Phys. Rev. Lett.* **108**, 227003 (2012)
- [121] T. Dahm, V. Hinkov, S.V. Borisenko, A.A. Kordyuk, V.B. Zabolotnyy, J. Fink, B. Büchner, D.J. Scalapino, W. Hanke, B. Keimer, *Nature Phys.* **5**, 217 (2009)
- [122] S.V. Borisenko, A.A. Kordyuk, V. Zabolotnyy, J. Geck, D. Inosov, A. Koitzsch, J. Fink, M. Knupfer, B. Büchner, V. Hinkov, C.T. Lin, B. Keimer, T. Wolf, S.G. Chiuzbaian, L. Patthey, R. Follath, *Phys. Rev. Lett.* **96**, 117004 (2006)
- [123] A.A. Kordyuk, S.V. Borisenko, V.B. Zabolotnyy, J. Geck, M. Knupfer, J. Fink, B. Büchner, C. T. Lin, B. Keimer, H. Berger, A.V. Pan, S. Komiya, Y. Ando, *Phys. Rev. Lett.* **97**, 017002 (2006)
- [124] J. Hwang, E.J. Nicol, T. Timusk, A. Knigavko, J.P. Carbotte, *Phys. Rev. Lett.* **98**, 207002 (2007)
- [125] E. van Heumen, E. Muhlethaler, A.B. Kuzmenko, H. Eisaki, W. Meevasana, M. Greven, and D. van der Marel, *Phys. Rev. B* **79**, 184512 (2009)
- [126] S. Dal Conte, C. Giannetti, G. Coslovich, F. Cilento, D. Bossini, T. Abebaw, F. Banfi, G. Ferrini, H. Eisaki, M. Greven, A. Damascelli, D. van der Marel, and F. Parmigiani, *Science* **335**, 1600 (2012)

- [127] A. Dubroka, M. Roessle, K.W. Kim, V.K. Malik, D. Munzar, D.N. Basov, A. Schafgans, S.J. Moon, C.T. Lin, D. Haug, V. Hinkov, B. Keimer, Th. Wolf, J.G. Storey, J.L. Tallon, C. Bernhard, *Phys. Rev. Lett.* **106**, 047006 (2011)
- [128] M.A. Metlitski, S. Sachdev, *Phys. Rev. B* **82**, 075128 (2010)
- [129] Y. Wang, A.V. Chubukov, *Phys. Rev. Lett.* **110**, 127001 (2013)
- [130] K.B. Efetov, H. Meier, C. Pépin, *Nat. Phys.* **9**, 442 (2013)
- [131] J.-H. Chung, T. Egami, R.J. McQueeney, M. Yethiraj, M. Arai, T. Yokoo, Y. Petrov, H.A. Mook, Y. Endoh, S. Tajima, C. Frost, F. Dogan, *Phys. Rev. B* **67**, 014517 (2003)
- [132] L. Pintschovius, D. Reznik, W. Reichardt, Y. Endoh, H. Hiraka, J.M. Tranquada, H. Uchiyama, T. Masui, S. Tajima, *Phys. Rev. B* **69**, 214506 (2004)
- [133] D. Reznik, L. Pintschovius, M. Ito, S. Iikubo, M. Sato, H. Goka, M. Fujita, K. Yamada, G.D. Gu, J.M. Tranquada, *Nature* **440**, 1170 (2006)
- [134] F. Stercel, T. Egami, H. A. Mook, M. Yethiraj, J.-H. Chung, M. Arai, C. Frost, F. Dogan, *Phys. Rev. B* **77**, 014502 (2008)
- [135] D. Reznik, L. Pintschovius, J.M. Tranquada, M. Arai, Y. Endoh, T. Masui, S. Tajima, *Phys. Rev. B* **78**, 094507 (2008)
- [136] B. Keimer, R.J. Birgeneau, A. Cassanho, Y. Endoh, M. Greven, M.A. Kastner, G. Shirane, *Z. Phys. B* **91**, 373 (1993)
- [137] V. Thampy, S. Blanco-Canosa, M. Garcia-Fernández, M.P.M. Dean, G.D. Gu, M. Först, T. Loew, B. Keimer, M. Le Tacon, S.B. Wilkins, J.P. Hill, *Phys. Rev. B* **88**, 024505 (2013)
- [138] For recent reviews, see P. Bourges, Y. Sidis, *Comptes Rendus Physique* **12**, 461 (2011); Y. Sidis, P. Bourges, [arXiv:1306.5124](https://arxiv.org/abs/1306.5124)
- [139] J. Mannhart, D. G. Schlom, *Science* **327**, 1607 (2010)
- [140] H.Y. Hwang, Y. Iwasa, M. Kawasaki, B. Keimer, N. Nagaosa, Y. Tokura, *Nature Mater.* **11**, 103 (2012)
- [141] A. Gozar, G. Logvenov, L. Fitting Kourkoutis, A.T. Bollinger, L.A. Giannuzzi, D.A. Muller, I. Bozovic, *Nature* **455**, 782 (2008)
- [142] S. Smadici, J.C.T. Lee, S. Wang, P. Abbamonte, G. Logvenov, A. Gozar, C. Deville Cavellin, I. Bozovic, *Phys. Rev. Lett.* **102**, 107004 (2009)
- [143] J. Chaloupka, G. Khaliullin, *Phys. Rev. Lett.* **100**, 016404 (2008)

- [144] A.V. Boris, Y. Matiks, E. Benckiser, A. Frano, P. Popovich, V. Hinkov, P. Wochner, M. Castro-Colin, E. Detemple, V.K. Malik, C. Bernhard, T. Prokscha, A. Suter, Z. Salman, E. Morenzoni, G. Cristiani, H.-U. Habermeier, B. Keimer, *Science* **332**, 937 (2011)
- [145] E. Benckiser, M.W. Haverkort, S. Brueck, E. Goering, S. Macke, A. Frano, X. Yang, O.K. Andersen, G. Cristiani, H.-U. Habermeier, A.V. Boris, I. Zegkinoglou, P. Wochner, H.-J. Kim, V. Hinkov, B. Keimer, *Nature Mater.* **10**, 189 (2011)
- [146] A. Frano, E. Schierle, M.W. Haverkort, Y. Lu, M. Wu, S. Blanco-Canosa, U. Nwankwo, A.V. Boris, P. Wochner, G. Cristiani, H.U. Habermeier, G. Logvenov, V. Hinkov, E. Benckiser, E. Weschke, B. Keimer, arXiv:1304.1469
- [147] D. Fausti, R.I. Tobey, N. Dean, S. Kaiser, A. Dienst, M.C. Hoffmann, S. Pyon, T. Takayama, H. Takagi, A. Cavalleri, *Science* **331**, 189 (2011)
- [148] S. Kaiser, D. Nicoletti, C.R. Hunt, W. Hu, I. Gierz, H.Y. Liu, M. Le Tacon, T. Loew, D. Haug, B. Keimer, A. Cavalleri, arXiv:1205.4661

10 Strongly Correlated Superconductivity

André-Marie S. Tremblay

Département de physique

Regroupement québécois sur les matériaux de pointe

Canadian Institute for Advanced Research

Sherbrooke, Québec J1K 2R1, Canada

Contents

1	Introduction	2
2	The Hubbard model	3
3	Weakly and strongly correlated antiferromagnets	4
3.1	Antiferromagnets: A qualitative discussion	5
3.2	Contrasting methods for weak and strong coupling antiferromagnets and their normal state	7
4	Weakly and strongly correlated superconductivity	9
4.1	Superconductors: A qualitative discussion	9
4.2	Contrasting methods for weakly and strongly correlated superconductors	11
5	High-temperature superconductors and organics: the view from dynamical mean-field theory	15
5.1	Quantum cluster approaches	16
5.2	Normal state and pseudogap	20
5.3	Superconducting state	24
6	Conclusion	30

1 Introduction

Band theory and the BCS-Eliashberg theory of superconductivity are arguably the most successful theories of condensed matter physics by the breadth and subtlety of the phenomena they explain. Experimental discoveries, however, clearly signal their failure in certain cases. Around 1940, it was discovered that some materials with an odd number of electrons per unit cell, for example NiO, were insulators instead of metals, a failure of band theory [1]. Peierls and Mott quickly realized that strong effective repulsion between electrons could explain this (Mott) insulating behaviour [2]. In 1979 and 1980, heavy fermion [3] and organic [4] superconductors were discovered, an apparent failure of BCS theory because the proximity of the superconducting phases to antiferromagnetism suggested the presence of strong electron-electron repulsion, contrary to the expected phonon-mediated attraction that gives rise to superconductivity in BCS. Superconductivity in the cuprates [5], in layered organic superconductors [6, 7], and in the pnictides [8] eventually followed the pattern: superconductivity appeared at the frontier of antiferromagnetism and, in the case of the layered organics, at the frontier of the Mott transition [9, 10], providing even more examples of superconductors falling outside the BCS paradigm [11, 12]. The materials that fall outside the range of applicability of band and of BCS theory are often called strongly correlated or quantum materials. They often exhibit spectacular properties, such as colossal magnetoresistance, giant thermopower, high-temperature superconductivity etc.

The failures of band theory and of the BCS-Eliashberg theory of superconductivity are in fact intimately related. In these lecture notes, we will be particularly concerned with the failure of BCS theory, and with the understanding of materials belonging to this category that we call strongly correlated superconductors. These superconductors have a normal state that is not a simple Fermi liquid and they exhibit surprising superconducting properties. For example, in the case of layered organic superconductors, they become better superconductors as the Mott transition to the insulating phase is approached [13].

These lecture notes are not a review article. The field is still evolving rapidly, even after more than 30 years of research. My aim is to provide for the student at this school an overview of the context and of some important concepts and results. I try to provide entries to the literature even for topics that are not discussed in detail here. Nevertheless, the reference list is far from exhaustive. An exhaustive list of all the references for just a few sub-topics would take more than the total number of pages I am allowed.

I will begin by introducing the one-band Hubbard model as the simplest model that contains the physics of interest, in particular the Mott transition. That model is 50 years old this year [14–16], yet it is far from fully understood. Section 3 will use antiferromagnetism as an example to introduce notions of weak and strong correlations and to contrast the theoretical methods that are used in both limits. Section 4 will do the same for superconductivity. Finally, Section 5 will explain some of the most recent results obtained with cluster generalizations of dynamical mean-field theory, approaches that allow one to explore the weak and strong correlation limits and the transition between both.

2 The Hubbard model

The one-band Hubbard model is given by

$$H = - \sum_{i,j,\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad (1)$$

where i and j label Wannier states on a lattice, $c_{i\sigma}^\dagger$ and $c_{i\sigma}$ are creation and annihilation operators for electrons of spin σ , $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ is the number of spin σ electrons on site i , $t_{ij} = t_{ji}^*$ are the hopping amplitudes, which can be taken as real in our case, and U is the on-site Coulomb repulsion. In general, we write t, t', t'' respectively for the first-, second-, and third-nearest neighbour hopping amplitudes.

This model is a drastic simplification of the complete many-body Hamiltonian, but we want to use it to understand the physics from the simplest point of view, without a large number of parameters. The first term of the Hubbard model Eq. (1) is diagonal in a momentum-space single-particle basis, the Bloch waves. There, the wave nature of the electron is manifest. If the interaction U is small compared to the bandwidth, perturbation theory and Fermi liquid theory hold [17, 18]. This is called the weak-coupling limit.

The interaction term in the Hubbard Hamiltonian, proportional to U , is diagonal in position space, i.e., in the Wannier orbital basis, exhibiting the particle nature of the electron. The motivation for that term is that once the interactions between electrons are screened, the dominant part of the interaction is on-site. Strong-coupling perturbation theory can be used if the bandwidth is small compared with the interaction [19–23].

Clearly, the intermediate-coupling limit will be most difficult, the electron exhibiting both wave and particle properties at once. The ground state will be entangled, i.e. very far from a product state of either Bloch (plane waves) or Wannier (localized) orbitals. We refer to materials in the strong or intermediate-coupling limits as strongly correlated.

When the interaction is the largest term and we are at half-filling, the solution of this Hamiltonian is a Mott insulating state. The ground state will be antiferromagnetic if there is not too much frustration. That can be seen as follows. If hopping vanishes, the ground state is 2^N -fold degenerate if there are N sites. Turning-on nearest-neighbour hopping, second order degenerate perturbation theory in t leads to an antiferromagnetic interaction $J \mathbf{S}_i \cdot \mathbf{S}_j$ with $J = 4t^2/U$ [24, 25]. This is the Heisenberg model. Since J is positive, this term will be smallest for anti-parallel spins. The energy is generally lowered in second-order perturbation theory. Parallel spins cannot lower their energy through this mechanism because the Pauli principle forbids the virtual, doubly occupied state. P.W. Anderson first proposed that the strong-coupling version of the Hubbard model could explain high-temperature superconductors [26].

A caricature of the difference between an ordinary band insulator and a Mott insulator is shown in Figure 1. Refer to the explanations in the caption.

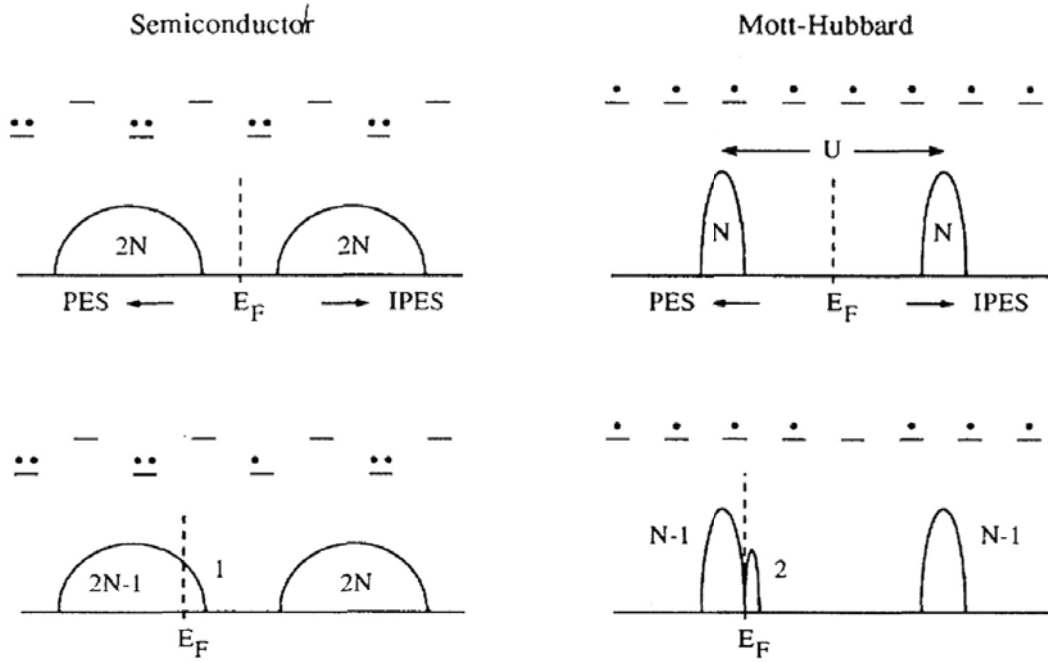


Fig. 1: Figure from Ref. [27]. In a band insulator, as illustrated on the top-left figure, the valence band is filled. For N sites on the lattice, there are $2N$ states in the valence band, the factor of 2 accounting for spins, and $2N$ states in the conduction band. PES on the figures refers to “Photoemission Spectrum” and IPES to “Inverse Photoemission Spectrum”. The small horizontal lines represent energy levels and the dots stand for electrons. In a Mott insulator, illustrated on the top-right figure, there are N states in the lower energy band (Lower Hubbard band) and N in the higher energy band (Upper Hubbard band), for a total of $2N$ as we expect in a single band. The two bands are separated by an energy U because if we add an electron to the already occupied states, it costs an energy U . Perhaps the most striking difference between a band and a Mott insulator manifests itself when the Fermi energy E_F is moved to dope the system with one hole. For the semiconductor, the Fermi energy moves, but the band does not rearrange itself. There is one unoccupied state right above the Fermi energy. This is seen on the bottom-left figure. On the bottom-right figure, we see that the situation is very different for a doped Mott insulator. With one electron missing, there are two states just above the Fermi energy, not one state only. Indeed one can add an electron with a spin up or down on the now unoccupied site. And only $N - 1$ states are left that will cost an additional energy U if we add an electron. Similarly, $N - 1$ states survive below the Fermi energy.

3 Weakly and strongly correlated antiferromagnets

A phase of matter is characterized by very general “emergent” properties, i.e., properties that are qualitatively different from those of constituent atoms [28–30]. For example, metals are shiny and they transport DC current. These are not properties of individual copper or gold atoms. It takes a finite amount of energy to excite these atoms from their ground state, so they cannot transport DC current. Also, their optical spectrum is made of discrete lines whereas metals reflect a continuous spectrum of light at low energy. In other words, the Fermi surface is an emergent property. Even in the presence of interactions, there is a jump in momentum

occupation number that defines the Fermi surface. This is the Landau Fermi liquid [17, 18]. Emergent properties appear at low energy, i.e., for excitation energies not far from the ground state. The same emergent properties arise from many different models. In the renormalization group language, phases are trivial fixed points, and many Hamiltonians flow to the same fixed point.

In this section, we use the antiferromagnetic phase to illustrate further what is meant by an emergent property and what properties of a phase depend qualitatively on whether we are dominated by band effects or by strong correlations. Theoretical methods appropriate for each limit are described in the last subsection.

3.1 Antiferromagnets: A qualitative discussion

Consider the nearest-neighbour Hubbard model at half-filling on the cubic lattice in three dimensions. At $T = 0$, there is a single phase, an antiferromagnet, whatever the value of the interaction U . One can increase U continuously without encountering a phase transition. There is an order parameter in the sense of Landau, in this case the staggered magnetization. This order parameter reflects the presence of a broken symmetry: time reversal, spin rotational symmetry and translation by a lattice spacing are broken, while time reversal accompanied by translation by a lattice spacing is preserved.

A single-particle gap and spin waves as Goldstone modes are emergent consequences of this broken symmetry. Despite the fact that we are in a single phase, there are qualitative differences between weak and strong coupling as soon as we probe higher energies. For example, at strong-coupling spin waves, at an energy scale of J , persist throughout the Brillouin zone, whereas at weak coupling they enter the particle-hole continuum and become Landau damped before we reach the zone boundary. The ordered moment is saturated to its maximum value when U is large enough but it can become arbitrarily small as U decreases.

The differences between weak and strong coupling are also striking at finite temperature. This is illustrated in a schematic fashion in Fig. 2a. The Néel temperature T_N , increases as we increase U because the instability of the normal state fundamentally comes from nesting. In other words, thinking again of perturbation theory, the flat parts of the Fermi surface are connected by the antiferromagnetic wave vector $Q = (\pi, \pi)$ which implies a vanishing energy denominator at that wave vector and thus large spin susceptibility χ with a phase transition occurring when $U\chi$ is large enough. At strong coupling, T_N decreases with increasing U because the spin stiffness is proportional to $J = 4t^2/U$. Since we can in principle vary the ratio t/U by changing pressure, it is clear that the pressure derivative of the Néel temperature has opposite sign at weak and strong coupling. The normal state is also very different. If we approach the transition from the left, as indicated by the arrow marked "Slater", we are in a metallic phase. We also say that the antiferromagnet that is born out of this metallic phase is an itinerant antiferromagnet. On the contrary, approaching the transition from the right, we come from an insulating (gapped) phase described by the Heisenberg model. Increasing U at fixed T above the maximum T_N , we have to cross over from a metal, that has a Fermi surface, to an insulator, that has local moments, a

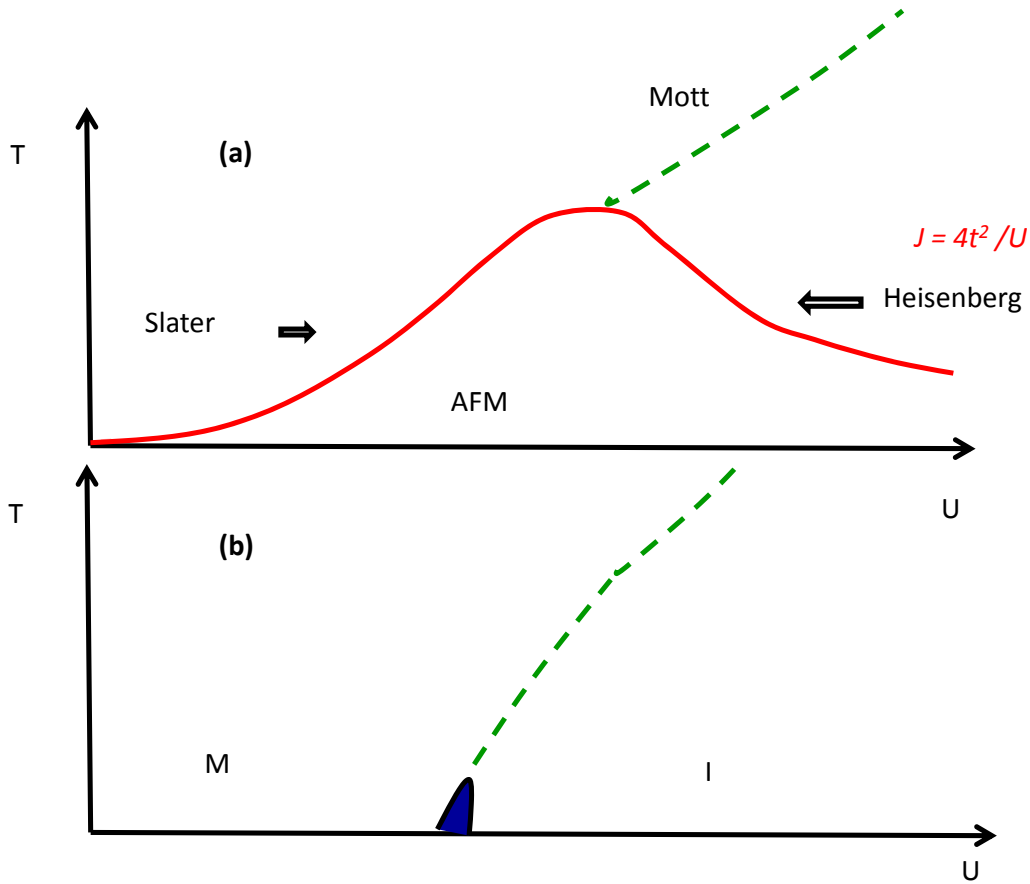


Fig. 2: Schematic phase diagram of the half-filled 3d Hubbard model in the temperature T vs. interaction U plane for perfect nesting. (a) The solid red line is the Néel temperature T_N below which the system is antiferromagnetic. Coming from the left, it is a metallic state that becomes unstable to antiferromagnetism by the Slater mechanism. Coming from the right, it is a gapped insulator with local moments described by the Heisenberg model that becomes unstable to antiferromagnetism. The dashed green line above the maximum T_N indicates a crossover from a metallic state with a Fermi surface to a gapped state with local moments. That crossover can be understood in (b) where antiferromagnetism is prevented from occurring. There dynamical mean-field theory predicts a first-order phase transition between a metal and a Mott insulator, with a coexistence region indicated in blue.

gap, and no Fermi surface. This highly non-trivial physics is indicated by a dashed green line in Fig. 2a and marked “Mott” [31].

Fig. 2b illustrates another way to understand the dashed green line. Imagine that antiferromagnetism does not occur before we reach zero temperature. This can be achieved in two dimensions where the Mermin-Wagner-Hohenberg [32,33] theorem prevents a broken continuous symmetry at finite temperature. More generally antiferromagnetism can be prevented by frustration [34]. Frustration can come either from longer-range hopping that leads to longer-range antiferromagnetic interactions or from the geometry of the lattice (e.g. the triangular lattice). In either case, it becomes impossible to minimize the energy of all the individual antiferromagnetic bonds without entering a contradiction. In Dynamical Mean-Field Theory [35–37], which we

will discuss later in more detail, antiferromagnetism can simply be prevented from occurring in the theory. In any case, what happens in the absence of antiferromagnetism is a first-order transition at $T = 0$ between a metal and an insulator. This is the Mott phase transition that ends at a critical temperature. This transition is seen, for example, in layered organic superconductors of the κ -BEDT family that we will briefly discuss later [6, 7]. The dashed green line at finite temperature is the crossover due to this transition. In the case where the antiferromagnetic phase is artificially prevented from occurring, the metallic and insulating phases at low temperature are metastable in the same way that the normal metal is a metastable state below the superconducting transition temperature. Just as it is useful to think of a Fermi liquid at zero temperature even when it is not the true ground state, it is useful to think of the zero-temperature Mott insulator even when it is a metastable state.

3.2 Contrasting methods for weak and strong coupling antiferromagnets and their normal state

In this subsection, I list some of the approaches that can be used to study the various limiting cases as well as their domain of applicability wherever possible. Note that the antiferromagnetic state can occur away from half-filling as well, so we also discuss states that would be best characterized as Fermi liquids.

3.2.1 Ordered state

The ordered state at weak coupling can be described for example by mean-field theory applied directly to the Hubbard model [38]. Considering spin waves as collective modes, one can proceed by analogy with phonons and compute the corresponding self-energy resulting from the exchange of spin waves. The staggered moment can then be obtained from the resulting Green function. It seems that this scheme interpolates smoothly and correctly from weak to strong coupling. More specifically, at very strong coupling in the $T = 0$ limit, the order parameter is renormalized down from its bare mean-field value by an amount very close to that predicted in a localized picture with a spin wave analysis of the Heisenberg model: In two dimensions on the square lattice, only two thirds of the full moment survives the zero-point fluctuations [39]. This is observed experimentally in the parent high-temperature superconductor La_2CuO_4 [40]. At strong coupling when the normal state is gapped, one can perform degenerate perturbation theory, or systematically apply canonical transformations to obtain an effective model [41, 42] that reduces to the Heisenberg model at very strong coupling. When the interaction is not strong enough, higher order corrections in t/U enter in the form of longer-range exchange interactions and so-called ring-exchanges [41–45]. Various methods such as $1/S$ expansions [46], $1/N$ expansions [47] or non-linear sigma models [48] are available.

Numerically, stochastic series expansions [49], high-temperature series expansions [50], quantum Monte Carlo (QMC) [51], world-line or worm algorithms [52, 53], and variational methods [54] are popular and accurate. Variational methods can be biased since one must guess a

wave function. Nevertheless, variational methods [55] and QMC have shown that in two dimensions, on the square lattice, the antiferromagnetic ground state is most likely [56]. States described by singlet formation at various length scales, so-called resonating valence bond spin liquids, are less stable. For introductions to various numerical methods, see the web archives of the summer schools [57] and [58].

3.2.2 The normal state

In strong coupling, the normal state is an insulator described mostly by the non-linear sigma model [59,60]. In the weak-coupling limit, the normal state is a metal described by Fermi liquid theory. To describe a normal state that can contain strong antiferromagnetic fluctuations [61], one needs to consider a version of Fermi liquid theory that holds on a lattice. Spin propagates in a diffusive manner. These collective modes are known as paramagnons. The instability of the normal state to antiferromagnetism can be studied by the random phase approximation (RPA), by self-consistent-renormalized theory (SCR) [62], by the fluctuation exchange approximation (FLEX) [63,64], by the functional renormalization group (FRG) [65–67], by field-theory methods [68,69], and by the two-particle self-consistent approach (TPSC) [70–72], to give some examples. Numerically, quantum Monte Carlo (QMC) is accurate and can serve as benchmark, but in many cases it cannot go to very low temperature because of the sign problem. That problem does not occur in the half-filled nearest-neighbor one-band Hubbard model, which can be studied at very low temperature with QMC [73].

The limitations of most of the above approaches have been discussed in the Appendices of Ref. [70]. Concerning TPSC, in short, it is non-perturbative and the most accurate of the analytical approaches at weak to intermediate coupling, as judged from benchmark QMC [70,72,74]. TPSC also satisfies the Pauli principle in the sense that the square of the occupation number for one spin species on a lattice site is equal to the occupation number itself, in other words $1^2 = 1$ and $0^2 = 0$. RPA is an example of a well known theory that violates this constraint (see, e.g., Appendix A3 of Ref. [70]). Also, TPSC satisfies conservation laws and a number of sum rules, including those which relate the spin susceptibility to the local moment and the charge susceptibility to the local charge. Most importantly, TPSC satisfies the Mermin-Wagner-Hohenberg theorem in two dimensions, contrary to RPA. Another effect included in TPSC is the renormalization of U coming from cross channels (Kanamori, Brückner screening) [15,75]. On a more technical level, TPSC does not assume a Migdal theorem in the calculation of the self-energy and the trace of the matrix product of the self-energy with the Green function satisfies the constraint that it is equal to twice the potential energy.

The most important prediction that came out of TPSC for the normal state on the two-dimensional square lattice, is that precursors of the antiferromagnetic ground state will occur when the antiferromagnetic correlation length becomes larger than the thermal de Broglie wave length $\hbar v_F / k_B T$ [76–78]. The latter is defined by the inverse of the wave vector spread that is caused by thermal excitations $\Delta \varepsilon \sim k_B T$. This result was verified experimentally [79] and it explains the pseudogap in electron-doped high-temperature superconductors [72,80,81].

4 Weakly and strongly correlated superconductivity

As discussed in the previous section, antiferromagnets have different properties depending on whether U is above or below the Mott transition, and appropriate theoretical methods must be chosen depending on the case. In this section, we discuss the analogous phenomenon for superconductivity. A priori, the superconducting state of a doped Mott insulator or a doped itinerant antiferromagnet are qualitatively different, even though some emergent properties are similar.

4.1 Superconductors: A qualitative discussion

As for antiferromagnets, the superconducting phase has emergent properties. For an s -wave superconductor, global charge conservation, or $U(1)$ symmetry, is broken. For a d -wave superconductor, in addition to breaking $U(1)$ symmetry, the order parameter does not transform trivially under rotation by $\pi/2$. It breaks the C_{4v} symmetry of the square lattice. In both cases we have singlet superconductivity, i.e., spin-rotational symmetry is preserved. In both cases, long-range forces push the Goldstone modes to the plasma frequency by the Anderson-Higgs mechanism [82]. The presence of symmetry-dictated nodes in the d -wave case is an emergent property with important experimental consequences: for example, the specific heat will vanish linearly with temperature and similarly for the thermal conductivity κ . The ratio κ/T reaches a universal constant in the $T = 0$ limit, i.e., that ratio is independent of disorder [83, 84]. The existence of a single-particle gap with nodes determined by symmetry is also an emergent property, but its detailed angular dependence and its size relative to other quantities, such as the transition temperature T_c , is dependent on details.

A possible source of confusion in terminology is that in the context of phonon mediated s -wave superconductivity, there is the notion of strong-coupling superconductivity. The word strong-coupling has a slightly different meaning from the one discussed up to now. The context should make it clear what we are discussing. Eliashberg theory describes phonon-mediated strong-coupling superconductivity [85–89]. In that case, quasiparticles survive the strong electron-phonon interaction, contrary to the case where strong electron-electron interactions destroy the quasiparticles in favour of local moments in the Mott insulator.

There are important quantitative differences between BCS and Eliashberg superconductors. In the latter case, the self-energy becomes frequency dependent so one can measure the effect of phonons on a frequency dependent gap function that influences in turn the tunnelling spectra. Predictions for the critical field $H_c(T)$ or for the ratio of the gap to T_c for example differ. The Eliashberg approach is the most accurate.

Let us return to our case, namely superconductors that arise from a doped Mott insulator (strongly correlated) and superconductors that arise from doping an itinerant antiferromagnet (weakly correlated). Again there are differences between both types of superconductors when we study more than just asymptotically small frequencies. For example, as we will see in Section 5, in strongly correlated superconductors, the gap is no-longer particle-hole sym-

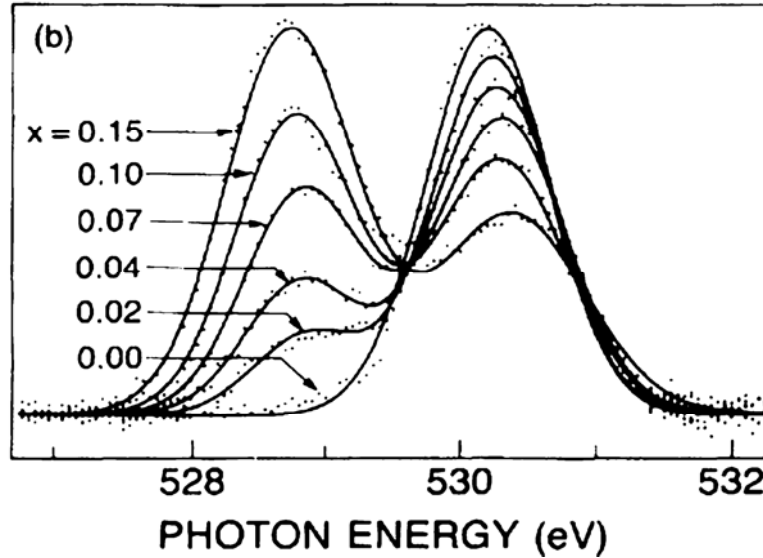


Fig. 3: Absorption spectrum for a sharp atomic level whose energy is about 0.5 keV below the Fermi level. At half-filling, or zero doping $x = 0$, it takes about 530 eV to excite an electron from the deep level to the upper Hubbard band. As one dopes, the upper Hubbard band still absorbs, but new states appear just above the Fermi level, allowing the X-ray to be absorbed at an energy about 2 eV below the upper Hubbard band. These states are illustrated on the bottom right panel of Fig. 1. The important point is that the spectral weight in the states just above the Fermi energy grows about twice as fast as the spectral weight in the upper Hubbard band decreases, in agreement with the cartoon picture of a doped Mott insulator. Figure from Ref. [102]

metric, and the transition temperature sometimes does not scale like the order parameter. A strongly-coupled superconductor is also more resilient to nearest-neighbor repulsion than a weakly-coupled one [90].

We should add a third category of strongly-correlated superconductors, namely superconductors that arise, in the context of the Hubbard model, at half-filling under a change of pressure. This is the case of the layered organics. There again superconductivity is very special since, contrary to naive expectations, it becomes stronger as we approach the Mott metal-insulator transition [13].

Just as for antiferromagnets, the normal state of weakly-correlated and of strongly-correlated superconductors is very different. Within the one-band Hubbard model as usual, the normal state of weakly-correlated superconductors is a Fermi liquid with antiferromagnetic fluctuations. In the case of strongly correlated superconductors, the normal state exhibits many strange properties, the most famous of which is probably the linear temperature dependence of resistivity that persists well above the Mott-Ioffe-Regel (MIR) limit [91–93]. This limit is defined as follows: Consider a simple Drude formula for the conductivity, $\sigma = ne^2\tau/m$, where n is the density, e is the electron charge, m the mass, and τ the collision time. Using the Fermi velocity v_F , one can convert the scattering time τ to a mean-free path ℓ . The MIR minimal conductivity is determined by stating that the mean-free path cannot be smaller than the Fermi wavelength. This means that, as a function of temperature, resistivity should saturate to that limit. This is

seen in detailed many-body calculations with TPSC [94] (These calculations do not include the possibility of the Mott transition). For a set of two-dimensional planes separated by a distance d , the MIR limit is set by $\hbar d/e^2$. That limit can be exceeded in doped Mott insulators [93,95]. The strange-metal properties, including the linear temperature dependence of the resistivity, are often considered as emergent properties of a new phase. Since new phases of matter are difficult to predict, one approach has been to find mean-field solutions of gauge-field theories. These gauge theories can be derived from Hubbard-Stratonovich transformations or from assumptions as to the nature of the emergent degrees of freedom [96].

As in the case of the antiferromagnet, the state above the optimal transition temperature in strongly correlated superconductors is a state where crossovers occur. There is much evidence that hole-doped high-temperature superconductors are doped Mott insulators. The high-temperature thermopower [97,98] and Hall coefficient [99] are examples of properties that are as expected from Mott insulators. We can verify the doped Mott insulator nature of the hole-doped cuprates from the experimental results for soft X-ray absorption spectroscopy as illustrated in Fig. 3. This figure should be compared with the cartoon in Fig. 1. More recent experimental results on this topic [100] and comparison with the Hubbard model [101] are available in the literature.

4.2 Contrasting methods for weakly and strongly correlated superconductors

In most phase transitions, a simple analysis at weak coupling indicates that the normal state is unstable towards a new phase. For example, one can compute an appropriate susceptibility in the normal state and observe that it can sometimes diverge at sufficiently low temperature, indicating an instability. For an antiferromagnet, we would compute the staggered spin-susceptibility. Alternatively, one could perform a mean-field factorization and verify that there is a self-consistent broken symmetry solution at low temperature. Neither of these two procedures however leads to a superconducting instability when we start from a purely repulsive Hubbard model. In this section, we will discuss how to overcome this.

Surprisingly, a mean-field factorization of the Hubbard model at strong coupling does lead to a superconducting d -wave ground state [103–105]. There are reasons, however, to doubt the approximations involved in a simple mean-field theory at strong coupling.

4.2.1 The normal state and its superconducting instability

In the weakly correlated case, the normal state must contain slow modes that replace the phonons to obtain a superconducting instability. It is as if the superconducting instability came at a second level of refinement of the theory. This all started with Kohn and Luttinger [106,107]: They noted that the interaction between two electrons is screened by the other electrons. Computing this screening to leading order, they found that in sufficiently high angular-momentum states and at sufficiently low temperature (usually very low) there is always a superconducting

instability in a Fermi liquid (see, e.g., Ref. [108] for a review). It is possible to do calculations in this spirit that are exact for infinitesimally small repulsive interactions [109].

Around 1986, before the discovery of high-temperature superconductivity, it was realized that if instead of a Fermi liquid in a continuum, one considers electrons on a lattice interacting with short-range repulsion, then the superconducting instability may occur more readily. More specifically, it was found that exchange of antiferromagnetic paramagnons between electrons can lead to a d -wave superconducting instability [110, 111]. This is somewhat analogous to what happens for ferromagnetic fluctuations in superfluid ^3He (for a recent exhaustive review of unconventional superconductivity, see Ref. [112]). These types of theories [113], like TPSC below, have some features that are qualitatively different from the BCS prediction [114, 115]. For example, the pairing symmetry depends more on the shape of the Fermi surface than on the single-particle density of states. Indeed, the shape of the Fermi surface determines the wave vector of the largest spin fluctuations, which in turn favor a given symmetry of the d -wave order parameter, for example $d_{x^2-y^2}$ vs. d_{xy} [116], depending on whether the antiferromagnetic fluctuations are in the (π, π) or $(\pi, 0)$ direction, respectively. It is also believed that conditions for maximal T_c are realized close to the quantum critical point of an apparently competing phase, such as an antiferromagnetic phase [117, 118], or a charge ordered phase [119].

In the case of quasi-one-dimensional conductors, a more satisfying way of obtaining d -wave superconductivity in the presence of repulsion consists of using the renormalization group [120, 121]. In this approach, one finds a succession of effective low-energy theories by eliminating perturbatively states that are far away from the Fermi surface. As more and more degrees of freedom are eliminated, i.e., as the cutoff decreases, the effective interactions for the low-energy theory can grow or decrease, or even change sign. In this approach then, all fluctuation channels are considered simultaneously, interfering with and influencing each other. The effective interaction in the particle-particle d -wave channel can become attractive, signalling a superconducting instability.

Whereas this approach is well justified in the one-dimensional and quasi one-dimensional cases from the logarithmic behavior of perturbation theory, in two dimensions more work is needed. Nevertheless, allowing the renormalized interactions to depend on all possible momenta, one can devise the so-called functional renormalization group. One can follow either the Wilson procedure [67], as was done originally for fermions by Bourbonnais [122, 123], or a functional approach [65, 66] closer in spirit to quantum field-theory approaches. d -wave superconductivity has been found in these approaches [124–126].

As we have already mentioned, in two dimensions, even at weak to intermediate coupling, the normal state out of which d -wave superconductivity emerges is not necessarily a simple Fermi liquid. It can have a pseudogap induced by antiferromagnetic fluctuations. Because TPSC is the only approach that can produce a pseudogap in two dimensions over a broad range of temperature, we mention some of the results of this approach [116, 127]. The approach is similar in spirit to the paramagnon theories above [110, 111, 113], but it is more appropriate because, as mentioned before, it satisfies the Mermin-Wagner theorem, is non-perturbative, and satisfies a number of sum rules. The irreducible vertex in the particle-particle channel is obtained from

functional derivatives. Here are some of the main results:

1. The pseudogap appears when the antiferromagnetic correlation length exceeds the thermal de Broglie wave length. This is why, even without competing order, T_c decreases as one approaches half-filling despite the fact that the AFM correlation length increases. This can be seen in Fig. 3 of Ref [127]. States are removed from the Fermi level and hence cannot lead to pairing. The dome is less pronounced when second-neighbor hopping t' is finite because the fraction of the Fermi surface where states are removed (hot spots) is smaller.
2. The superconducting T_c depends rather strongly on t' . At fixed filling, there is an optimal frustration, namely a value of t' where T_c is maximum as illustrated in Fig. 5 of Ref. [116].
3. Fig. 6 of Ref. [116] shows that T_c can occur below the pseudogap temperature or above. (The caption should read $U = 6$ instead of $U = 4$). For the cases considered, that include optimal frustration, the AFM correlation length at the maximum T_c is about 9 lattice spacings, as in Ref. [128]. Elsewhere, it takes larger values at T_c .
4. A correlation between resistivity and T_c in the pnictides, the electron-doped cuprates, and the quasi one-dimensional organics was well established experimentally in Ref. [129]. Theoretically it is well understood for the quasi one-dimensional organics [121]. (It seems to be satisfied in TPSC, as illustrated by Fig. 5 of Ref. [81], but the analytical continuation has some uncertainties.)
5. There is strong evidence that the electron-doped cuprates provide an example where antiferromagnetically mediated superconductivity can be verified. Figs. 3 and 4 of the paper by the Greven group [79] show that at optimal T_c the AFM correlation length is of the order of 10 lattice spacings. The photoemission spectrum and the AFM correlation length obtained from the Hubbard model [80] with $t' = -0.175t$, $t'' = 0.05t$ and $U = 6.25t$ agree with experiment. In particular, in TPSC one obtains the dynamical exponent $z = 1$ at the antiferromagnetic quantum critical point [81], as observed in the experiments [79]. This comes from the fact that at that quantum critical point, the Fermi surface touches only one point when it crosses the antiferromagnetic zone boundary. The strange discontinuous doping dependence of the AFM correlation length near the optimal T_c obtained by Greven's group is however unexplained. The important interference between antiferromagnetism and d -wave superconductivity found with the functional renormalization group [121] is not included in TPSC calculations however. I find that in the cuprate family, the electron-doped systems are those for which the case for a quantum-critical scenario [117, 118, 130, 131] for superconductivity is the most justified.

In a doped Mott insulator, the problem becomes very difficult. The normal state is expected to be very anomalous, as we have discussed above. Based on the idea of emergent behavior, many researchers have considered slave-particle approaches [96]. The exact creation operators in these approaches are represented in a larger Hilbert space by products of fermionic and

bosonic degrees of freedom with constraints that restrict the theory to the original Hilbert space. There is large variety of these approaches: Slave bosons of various kinds [132–134], slave fermions [135,136], slave rotors [137], slave spins [138], or other field theory approaches [139]. Depending on which of these methods is used, one obtains a different kind of mean-field theory with gauge fields that are used to enforce relaxed versions of the exact constraints that the theories should satisfy. Since there are many possible mean-field theories for the same starting Hamiltonian that give different answers, and lacking variational principle to decide between them [140], one must rely on intuition and on a strong belief of emergence in this kind of approaches.

At strong coupling, the pseudogap in the normal state can also be treated phenomenologically quite successfully with the Yang Rice Zhang (YRZ) model [141]. Inspired by renormalized mean-field theory, which I discuss briefly in the following section, this approach suggests a model for the Green function that can then be used to compute many observable properties [142, 143].

The normal state may also be treated by numerical methods, such as variational approaches, or by quantum cluster approaches. Section 5 below is devoted to this methodology.

It should be pointed out that near the Mott transition at $U = 6$ on the square lattice where TPSC ceases to be valid, the value of optimal T_c that is found in Ref. [127] is close to that found with the quantum cluster approaches discussed in Sec. 5 below [144]. The same statement is valid at $U = 4$ [74, 127, 145], where this time the quantum cluster approaches are less accurate than at larger U . This agreement of non-perturbative weak and strong coupling methods at intermediate coupling gives us confidence in the validity of the results.

4.2.2 Ordered state

Whereas the Hubbard model does not have a simple mean-field d -wave solution, its strong-coupling version, namely the t - J model does [103–105]. More specifically if we perform second-order degenerate perturbation theory starting from the large U limit, the effective low-energy Hamiltonian reduces to

$$H = - \sum_{i,j,\sigma} t_{ij} P c_{i\sigma}^\dagger c_{j\sigma} P + J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (2)$$

where P is a projection operator ensuring that hopping does not lead to double occupancies. In the above expression, correlated hopping terms and density-density terms have been neglected. To find superconductivity, one proceeds like Anderson [26] and writes the spin operators in terms of Pauli matrices $\vec{\sigma}$ and creation-annihilation operators so that the Hamiltonian reduces to

$$H = - \sum_{i,j,\sigma} t_{ij} P c_{i\sigma}^\dagger c_{j\sigma} P + J \sum_{\langle i,j \rangle, \alpha, \beta, \gamma, \delta} \left(\frac{1}{2} c_{i\alpha}^\dagger \vec{\sigma}_{\alpha\beta} c_{i\beta} \right) \cdot \left(\frac{1}{2} c_{j\gamma}^\dagger \vec{\sigma}_{\gamma\delta} c_{j\delta} \right). \quad (3)$$

Defining the d -wave order parameter, with N the number of sites and unit lattice spacing, as

$$d = \langle \hat{d} \rangle = \frac{1}{N} \sum_{\mathbf{k}} (\cos k_x - \cos k_y) \langle c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow} \rangle, \quad (4)$$

a mean-field factorization, including the possibility of Néel order m , leads to the mean-field Hamiltonian

$$H_{MF} = \sum_{\mathbf{k}, \sigma} \varepsilon(\mathbf{k}) c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} - 4Jm\hat{m} - Jd(\hat{d} + \hat{d}^\dagger). \quad (5)$$

The dispersion relation $\varepsilon(\mathbf{k})$ is obtained by replacing the projection operators by the average doping. The d -wave nature of the order was suggested in Refs. [104, 146]. The superconducting state in this approach is not much different from an ordinary BCS superconductor, but with renormalized hopping parameters. In the above approach, it is clear that the instantaneous interaction J causes the binding. This has led Anderson to doubt the existence of a “pairing glue” in strongly correlated superconductors [147]. We will see in the following section that more detailed numerical calculations give a different perspective [148, 149].

The intuitive weak-coupling argument for the existence of a d -wave superconductor in the presence of antiferromagnetic fluctuations [115, 150] starts from the BCS gap equation

$$\Delta_{\mathbf{p}} = - \int \frac{d\mathbf{p}'}{(2\pi)^2} U(\mathbf{p} - \mathbf{p}') \frac{\Delta_{\mathbf{p}'}}{2E_{\mathbf{p}'}} (1 - 2f(E_{\mathbf{p}'})) \quad (6)$$

where $E_{\mathbf{p}} = \sqrt{\varepsilon_{\mathbf{p}}^2 + \Delta_{\mathbf{p}}^2}$ and f is the Fermi function. In the case of an s -wave superconductor, the gap is independent of \mathbf{p} so it can be simplified on both sides of the equation. There will be a solution only if U is negative since all other factors on the right-hand side are positive. In the presence of a repulsive interaction, in other words when U is positive, a solution where the order parameter changes sign is possible. For example, suppose \mathbf{p} on the left-hand side is in the $(\pi, 0)$ direction of a square lattice. Then if, because of antiferromagnetic fluctuations, $U(\mathbf{p} - \mathbf{p}')$ is peaked near (π, π) , then the most important contributions to the integral come from points such that \mathbf{p}' is near $(0, \pi)$ or $(\pi, 0)$, where the gap has a different sign. That sign will cancel with the overall minus sign on the right-hand side, making a solution possible.

Superconductivity has also been studied with many strong-coupling methods, including the slave-particle-gauge-theory approaches [96], the composite-operator method [151], and the YRZ approach mentioned above [141]. In the next section, we focus on quantum cluster approaches.

5 High-temperature superconductors and organics: the view from dynamical mean-field theory

In the presence of a Mott transition, the unbiased numerical method of choice is dynamical mean-field theory. When generalized to a cluster [152–154], one sometimes refers to these methods as “quantum cluster approaches”. For reviews, see Refs. [74, 155, 156]. The advantage of this method is that all short-range dynamical and spatial correlations are included. Long range spatial correlations on the other hand are included at the mean-field level as broken symmetry states. The symmetry is broken in the bath only, not on the cluster. Long-wavelength particle-hole and particle-particle fluctuations are, however, missing.

After a short formal derivation of the method, we will present a few results for the normal and for the superconducting state. In both cases, we will emphasize the new physics that arises in the strong coupling regime.

5.1 Quantum cluster approaches

In short, dynamical mean-field theory (DMFT) can be understood simply as follows: In infinite dimension one can show that the self-energy depends only on frequency [157]. To solve the problem exactly, one considers a single site with a Hubbard interaction immersed in a bath of non-interacting electrons [35–37]. Solving this problem, one obtains a self-energy that should be that entering the full lattice Green function. The bath is determined self-consistently by requiring that when the lattice Green function is projected on a single site, one obtains the same Green function as that of the single-site in a bath problem. In practice, this approach works well in three dimensions. In lower dimension, the self-energy acquires a momentum dependence and one must immerse a small interacting cluster in a self-consistent bath. One usually refers to the cluster or the single-site as “the impurity”. The rest of this subsection is adapted from Ref. [74]; for a more detailed derivation see the lecture of R. Eder. It is not necessary to understand the details of this derivation to follow the rest of the lecture notes.

Formally, the self-energy functional approach, devised by Potthoff [158–161], allows one to consider various cluster schemes from a unified point of view. It begins with $\Omega_t[G]$, a functional of the Green function

$$\Omega_t[G] = \Phi[G] - \text{Tr}((G_{0t}^{-1} - G^{-1})G) + \text{Tr} \ln(-G). \quad (7)$$

The Luttinger Ward functional $\Phi[G]$ entering this equation is the sum of two-particle irreducible skeleton diagrams. For our purposes, what is important is that (i) The functional derivative of $\Phi[G]$ is the self-energy

$$\frac{\delta \Phi[G]}{\delta G} = \Sigma \quad (8)$$

and (ii) it is a universal functional of G in the following sense: whatever the form of the one-body Hamiltonian, it depends only on the interaction and, functionally, it depends only on G and on the interaction, not on the one-body Hamiltonian. The dependence of the functional $\Omega_t[G]$ on the one-body part of the Hamiltonian is denoted by the subscript t and it comes only through G_{0t}^{-1} appearing on the right-hand side of Eq. (7).

The functional $\Omega_t[G]$ has the important property that it is stationary when G takes the value prescribed by Dyson’s equation. Indeed, given the last two equations, the Euler equation takes the form

$$\frac{\delta \Omega_t[G]}{\delta G} = \Sigma - G_{0t}^{-1} + G^{-1} = 0. \quad (9)$$

This is a dynamic variational principle since it involves the frequency appearing in the Green function, in other words excited states are involved in the variation. At this stationary point, and only there, $\Omega_t[G]$ is equal to the grand potential. Contrary to Ritz’s variational principle, this

variation does not tell us whether the stationary point of $\Omega_t[G]$ is a minimum, a maximum, or a saddle point.

Suppose we can locally invert Eq. (8) for the self-energy to write G as a functional of Σ . We can use this result to write,

$$\Omega_t[\Sigma] = F[\Sigma] - \text{Tr} \ln(-G_{0t}^{-1} + \Sigma), \quad (10)$$

where we defined

$$F[\Sigma] = \Phi[G] - \text{Tr}(\Sigma G) \quad (11)$$

and where it is implicit that $G = G[\Sigma]$ is now a functional of Σ . We refer to this functional as the Potthoff functional. Potthoff called this method the self-energy functional approach. Several types of quantum cluster approaches may be derived from this functional. A crucial observation is that $F[\Sigma]$, along with the expression (8) for the derivative of the Luttinger-Ward functional, define the Legendre transform of the Luttinger-Ward functional. It is easy to verify that

$$\frac{\delta F[\Sigma]}{\delta \Sigma} = \frac{\delta \Phi[G]}{\delta G} \frac{\delta G[\Sigma]}{\delta \Sigma} - \Sigma \frac{\delta G[\Sigma]}{\delta \Sigma} - G = -G. \quad (12)$$

Hence, $\Omega_t[\Sigma]$ is stationary with respect to Σ when Dyson's equation is satisfied

$$\frac{\delta \Omega_t[\Sigma]}{\delta \Sigma} = -G + (G_{0t}^{-1} - \Sigma)^{-1} = 0. \quad (13)$$

We now take advantage of the fact that $F[\Sigma]$ is universal, i.e., that it depends only on the interaction part of the Hamiltonian and not on the one-body part. This follows from the universal character of its Legendre transform $\Phi[G]$. We thus evaluate $F[\Sigma]$ exactly for a Hamiltonian H' that shares the same interaction part as the Hubbard Hamiltonian, but that is exactly solvable. This Hamiltonian H' is taken as a cluster decomposition of the original problem, i.e., we tile the infinite lattice into identical, disconnected clusters that can be solved exactly. Denoting the corresponding quantities with a prime, we obtain,

$$\Omega_{t'}[\Sigma'] = F[\Sigma'] - \text{Tr} \ln(-G_{0t'}^{-1} + \Sigma') \quad (14)$$

from which we can extract $F[\Sigma']$. It follows that

$$\Omega_t[\Sigma'] = \Omega_{t'}[\Sigma'] + \text{Tr} \ln(-G_{0t}^{-1} + \Sigma') - \text{Tr} \ln(-G_{0t'}^{-1} + \Sigma'). \quad (15)$$

The fact that the self-energy (real and imaginary parts) Σ' is restricted to the exact self-energy of the cluster problem H' , means that variational parameters appear in the definition of the one-body part of H' .

In practice, we look for values of the cluster one-body parameters t' such that $\delta \Omega_t[\Sigma'] / \delta t' = 0$. It is useful for what follows to write the latter equation formally, although we do not use it in actual calculations. Given that $\Omega_{t'}[\Sigma']$ is the grand potential evaluated for the cluster,

$\partial\Omega_{\mathbf{t}'}[\Sigma']/\partial\mathbf{t}'$ is cancelled by the explicit \mathbf{t}' dependence of $\text{Tr} \ln(-G_{0\mathbf{t}'}^{-1} + \Sigma')$ and we are left with

$$0 = \frac{\delta\Omega_{\mathbf{t}'}[\Sigma']}{\delta\Sigma'} \frac{\delta\Sigma'}{\delta\mathbf{t}'} = -\text{Tr} \left[\left(\frac{1}{G_{0\mathbf{t}'}^{-1} - \Sigma'} - \frac{1}{G_{0\mathbf{t}}^{-1} - \Sigma'} \right) \frac{\delta\Sigma'}{\delta\mathbf{t}'} \right]. \quad (16)$$

Given that the clusters corresponding to \mathbf{t}' are disconnected and that translation symmetry holds on the superlattice of clusters, each of which contains N_c sites, the last equation may be written

$$\sum_{\omega_n} \sum_{\mu\nu} \left[\frac{N}{N_c} \left(\frac{1}{G_{0\mathbf{t}'}^{-1} - \Sigma'(i\omega_n)} \right)_{\mu\nu} - \sum_{\tilde{\mathbf{k}}} \left(\frac{1}{G_{0\mathbf{t}}^{-1}(\tilde{\mathbf{k}}) - \Sigma'(i\omega_n)} \right)_{\mu\nu} \right] \frac{\delta\Sigma'_{\nu\mu}(i\omega_n)}{\delta\mathbf{t}'} = 0. \quad (17)$$

5.1.1 Cellular dynamical mean-field theory

The Cellular dynamical mean-field theory (CDMFT) [153] is obtained by including in the cluster Hamiltonian H' a bath of uncorrelated electrons that somehow must mimic the effect of the rest on the lattice on the cluster. Explicitly, H' takes the form

$$H' = - \sum_{\mu,\nu,\sigma} t'_{\mu\nu} c_{\mu\sigma}^\dagger c_{\nu\sigma} + U \sum_{\mu} n_{\mu\uparrow} n_{\mu\downarrow} a_{\alpha\sigma} + \text{H.c.} + \sum_{\alpha} \epsilon_{\alpha} a_{\alpha\sigma}^\dagger a_{\alpha\sigma} \quad (18)$$

where $a_{\alpha\sigma}$ annihilates an electron of spin σ on a bath orbital labeled α . The bath is characterized by the energy ϵ_{α} of each orbital and the bath-cluster hybridization matrix $V_{\mu\alpha}$. The effect of the bath on the electron Green function is encapsulated in the so-called hybridization function

$$\Gamma_{\mu\nu}(\omega) = \sum_{\alpha} \frac{V_{\mu\alpha} V_{\nu\alpha}^*}{\omega - \epsilon_{\alpha}} \quad (19)$$

which enters the Green function as

$$[G'^{-1}]_{\mu\nu} = \omega + \mu - t'_{\mu\nu} - \Gamma_{\mu\nu}(\omega) - \Sigma_{\mu\nu}(\omega). \quad (20)$$

Moreover, the CDMFT does not look for a strict solution of the Euler equation (17), but tries instead to set each of the terms between brackets to zero separately. Since the Euler equation (17) can be seen as a scalar product, CDMFT requires that the modulus of one of the vectors vanish to make the scalar product vanish. From a heuristic point of view, it is as if each component of the Green function in the cluster were equal to the corresponding component deduced from the lattice Green function. This clearly reduces to single site DMFT when there is only one lattice site.

Clearly, in this approach we have lost translational invariance. The self-energy and Green functions depends not only on the superlattice wave vector $\tilde{\mathbf{k}}$, but also on cluster indices. By going to Fourier space labeled by as many \mathbf{K} values as cluster indices, the self-energy or the Green function may be written as functions of two momenta, for example $G(\tilde{\mathbf{k}} + \mathbf{K}, \tilde{\mathbf{k}} + \mathbf{K}')$. In the jargon, periodizing a function to recover translational-invariance, corresponds to keeping only the diagonal pieces, $\mathbf{K} = \mathbf{K}'$. The final lattice Green function from which one computes

observable quantities is obtained by periodizing the self-energy [153], the cumulants [162], or the Green function itself. The last approach can be justified within the self-energy functional mentioned above because it corresponds to the Green function needed to obtain the density from $\partial\Omega/\partial\mu = -\text{Tr}(G)$. Periodization of the self-energy gives additional unphysical states in the Mott gap [163]. There exists also a version of DMFT formulated in terms of cumulants Ref. [164]. The fact that the cumulants are maximally local is often used to justify their periodization [162]. Explicit comparisons of all three methods appear in Ref. [165].

The DCA [152] cannot be formulated within the self-energy functional approach [166]. It is based on the idea of discretizing irreducible quantities, such as the self-energy, in reciprocal space. It is believed to converge faster for $\mathbf{q} = \mathbf{0}$ quantities, whereas CDMFT converges exponentially fast for local quantities [167–170].

5.1.2 Impurity solver

The problem of a cluster in a bath of non-interacting electrons is not trivial. It can be attacked by a variety of methods, ranging from exact diagonalization [171–177] and numerical renormalization group [178] to Quantum Monte Carlo [152]. The continuous-time quantum-Monte-Carlo solver can handle an infinite bath and is the only one that is in principle exact, apart from controllable statistical uncertainties [179].

For illustration, I briefly discuss the exact diagonalization solver introduced in Ref. [171] in the context of DMFT (i.e., a single site). For a pedagogical introduction, see also [180,181]. When the bath is discretized, i.e., is made of a finite number of bath “orbitals”, the left-hand side of Eq. (17) cannot vanish separately for each frequency, since the number of degrees of freedom in the bath is insufficient. Instead, one adopts the following self-consistent scheme: (i) start with a guess value of the bath parameters ($V_{\mu\alpha}, \epsilon_\alpha$) and solve the cluster Hamiltonian H' numerically; (ii) then calculate the combination

$$\hat{\mathcal{G}}_0^{-1} = \left[\sum_{\mathbf{k}} \frac{1}{\hat{G}_{0t}^{-1}(\mathbf{k}) - \hat{\Sigma}'(i\omega_n)} \right]^{-1} + \hat{\Sigma}'(i\omega_n) \quad (21)$$

and (iii) minimize the following canonically invariant distance function

$$d = \sum_{n,\mu,\nu} \left| \left(i\omega_n + \mu - \hat{t}' - \hat{F}(i\omega_n) - \hat{\mathcal{G}}_0^{-1} \right)_{\mu\nu} \right|^2 \quad (22)$$

over the set of bath parameters (changing the bath parameters at this step does not require a new solution of the Hamiltonian H' , but merely a recalculation of the hybridization function \hat{F}). The bath parameters obtained from this minimization are then put back into step (i) and the procedure is iterated until convergence.

In practice, the distance function (22) can take various forms, for instance by adding a frequency-dependent weight in order to emphasize low-frequency properties [172, 182, 183] or by using a sharp frequency cutoff [172]. These weighting factors can be considered as rough approximations for the missing factor $\delta\Sigma'_{\nu\mu}(i\omega_n)/\delta t'$ in the Euler equation (17). The frequencies are

summed over a discrete, regular grid along the imaginary axis, defined by some fictitious inverse temperature β , typically of the order of 20 or 40 (in units of t^{-1}).

5.2 Normal state and pseudogap

Close to half-filling, as we discussed above, the normal state of high-temperature superconductors exhibits special properties. Up to optimal doping, roughly, there is a doping dependent temperature T^* , where a gap slowly opens up as temperature is decreased. This phenomenon is called a "pseudogap". We have discussed it briefly above. T^* decreases monotonically with increasing doping. The signature of the pseudogap is seen in many physical properties. For example, the uniform magnetic spin susceptibility, measured by the Knight shift in nuclear magnetic resonance [184], decreases strongly with temperature, contrary to an ordinary metal, where the spin susceptibility, also known as Pauli susceptibility, is temperature independent. Also, the single-particle density of states develops a dip between two energies on either side of the Fermi energy whose separation is almost temperature independent [185]. Angle-Resolved-Photoemission (ARPES) shows that states are pushed away from the Fermi energy in certain directions [186, 187]. To end this non-exhaustive list, we mention that the c -axis resistivity increases with decreasing temperature while the optical conductivity develops a pseudogap [188]. For a short review, see Ref. [189]. An older review appears in Ref. [190]. There are three broad classes of mechanisms for opening a pseudogap:

1. Since phase transitions often open up gaps, the pseudogap could appear because of a first-order transition rounded by disorder.
2. In two-dimensions the Mermin-Wagner-Hohenberg theorem prohibits the breaking of a continuous symmetry. However, there is a regime with strong fluctuations that leads to the opening of a precursor of the true gap that will appear in the zero-temperature ordered state. We briefly explained this mechanism at the end of Sec. 3.2.2 and its application to electron-doped cuprates, which are less strongly coupled than the hole-doped ones [191, 192]. Further details are in Ref. [72].
3. Mott physics by itself can lead to a pseudogap. This mechanism, different from the previous ones, as emphasized before [191, 193], is considered in the present section. As discussed at the end of Section 4.1 and in Fig. 3, the hole-doped cuprates are doped Mott insulators, so this last possibility for a pseudogap needs to be investigated.

Before proceeding further, note that the candidates for the order parameter of a phase transition associated with the pseudogap are numerous: stripes [203], nematic order [200], d -density wave [204], antiferromagnetism [205], There is strong evidence in several cuprates of a charge-density wave [196, 198, 206], anticipated from transport [207] and quantum oscillations [208], and of intra-unit cell nematic order [209]. All of this is accompanied by Fermi surface reconstruction [208, 210, 211]. Time-reversal symmetry breaking also occurs, as evidenced by the Kerr effect [195] and by the existence of intra unit-cell spontaneous currents

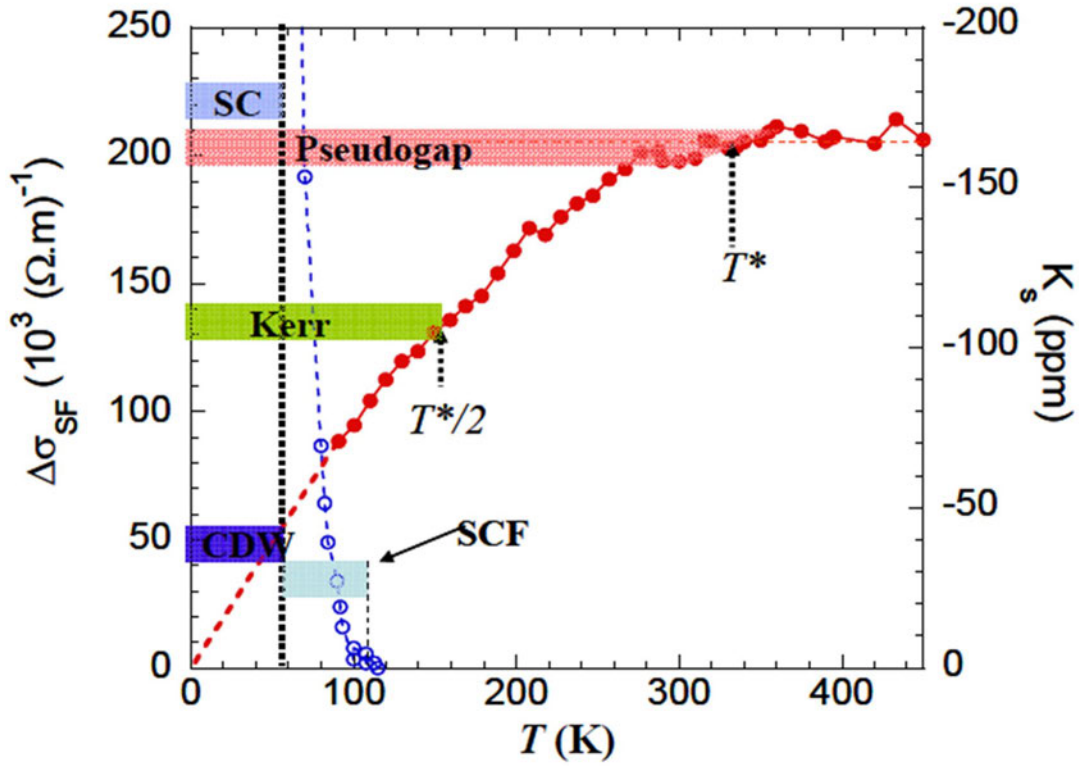


Fig. 4: In red, the experimental spin contribution to the Knight shift (on the right vertical axis) for $\text{YBCO}_{6.6}$. [194]. The illustrated phase transitions for superconducting transition temperature, for the Kerr signal [195], and for the high-field CDW from NQR quadrupole effects [196] and ultrasound velocity data [197]. The CDW transition from X-Ray diffraction in zero field [198] coincides with the onset of the Kerr effect. Recent resonant ultrasound measurements, suggest a true phase transition at T^* [199], like the Nernst signal [200]. The sharp drop in the superconducting fluctuation conductivity (SCF) [201] is also illustrated in blue with the corresponding vertical axis on the left. Figure from Ref. [202].

as evidenced by polarized neutron scattering [212, 213]. Nevertheless, it seems that such order appear at lower temperature than the pseudogap temperature [202]. The ordering seems a consequence rather than the cause of the pseudogap [214]. This is illustrated by Fig. 4 and discussed further in the corresponding caption. Clearly, some of these orders, such as intra-unit cell spontaneous currents [212, 213], cannot be explained within a one-band model. Nevertheless, since the order generally appears below the T^* illustrated in the figure, see, however, the caption of Fig. 4, it is worth investigating the predictions of the simple one-band model.

Early finite temperature DCA [215], and zero-temperature exact diagonalizations with cluster perturbation theory [191] and with CDMFT [172, 175, 176, 216, 217], have shown that the calculated pseudogap for ARPES is very close to experiment. Several recent calculations with CDMFT [216, 218, 219] or DCA [218, 220] using continuous-time quantum Monte Carlo at finite temperature as an impurity solver have found similar results.

Fig. 5 illustrates the essential features of the normal state phase diagram for the Hubbard model with nearest-neighbor hopping only on a 2×2 plaquette [221, 222]. When the doping δ vanishes

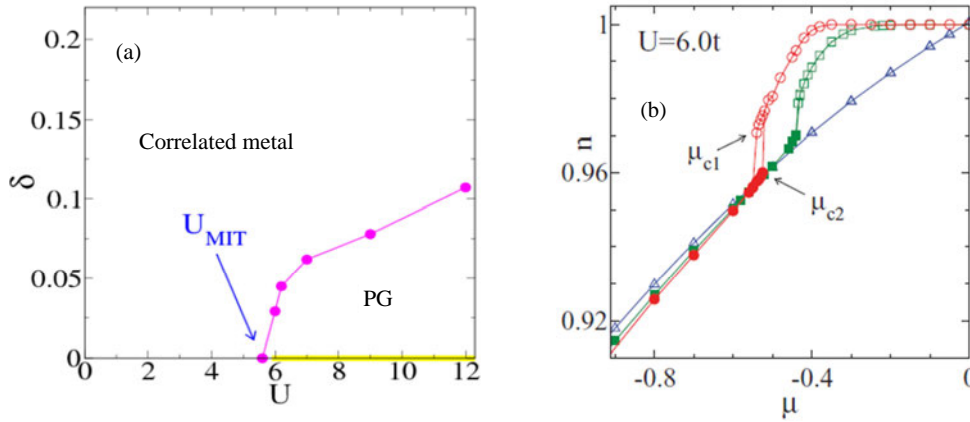


Fig. 5: (a) Zero-temperature extrapolation of the normal-state phase diagram for a 2×2 plaquette for the Hubbard model with nearest-neighbor hopping. In addition to the Mott transition at zero doping where the insulating phase is in yellow, there is a first-order transition line that separates a phase with a pseudogap from a strongly correlated metal. Figure adapted by G. Sordi from [222]. Signs of a first order transition have also been seen in Ref. [176]. (b) Filling n versus chemical potential μ for different temperatures: At high temperature, $T = 1/10$, there is a single curve represented by the blue line with triangles. At lower temperatures, $T = 1/25$ in green with squares, and at $T = 1/50$ in red with circles, there are clear signs of a first order transition [221, 222].

in Fig. 5(a), the insulating phase, represented by the yellow line, begins around $U = 5.8$. Hysteresis (not shown) occurs as a function of U . From this line, emerges another first-order transition line that separates two types of metals: A metallic state with a pseudogap near half-filling, and a correlated metal away from it. The transition between the two metals is well illustrated in Fig. 5(b). Consider the $n(\mu)$ curve at $T = 1/50$ represented by circles on a red line. Decreasing μ from $\mu = 0$, the density remains fixed at $n = 1$ for a relatively large range of μ because the Mott gap is opened and the chemical potential is in the gap. Around $\mu = -0.4$ one enters a compressible phase, i.e., $dn/d\mu$ is finite. The rounded crossover is due to the finite temperature. It should become a discontinuous change in slope at $T = 0$. The jump in filling and the hysteresis is obvious near $\mu = -0.55$.

Let us now fix the interaction strength to $U = 6.2$ and look at the phase diagram in the doping-temperature plane in Fig. 6. This is obtained for the normal state, without allowing for antiferromagnetism [224] (Disregard for the moment the superconducting region in blue. It will be discussed in the next subsection.¹) The first-order transition is illustrated by the shaded region between the red lines. Various crossovers are identified as described in the caption. Let us focus on the two crossover lines associated with the uniform magnetic susceptibility. The purple solid line with circles that appears towards the top of the graph identifies, for a given doping, the temperature at which the susceptibility starts to fall just after it reaches a broad maximum. This was identified as T^* in Fig. 4.² The purple line ends before it reaches zero temperature

¹Recent progress on the ergodicity of the hybridization expansion for continuous-time quantum Monte Carlo leads to results for the superconducting phase that are qualitatively similar but quantitatively different from those that appear for this phase in Figs. 6 and 7(c). These results will appear later. [144].

²Note that T^* in Fig. 6 refers to the inflection point of the zero frequency density of states as a function of

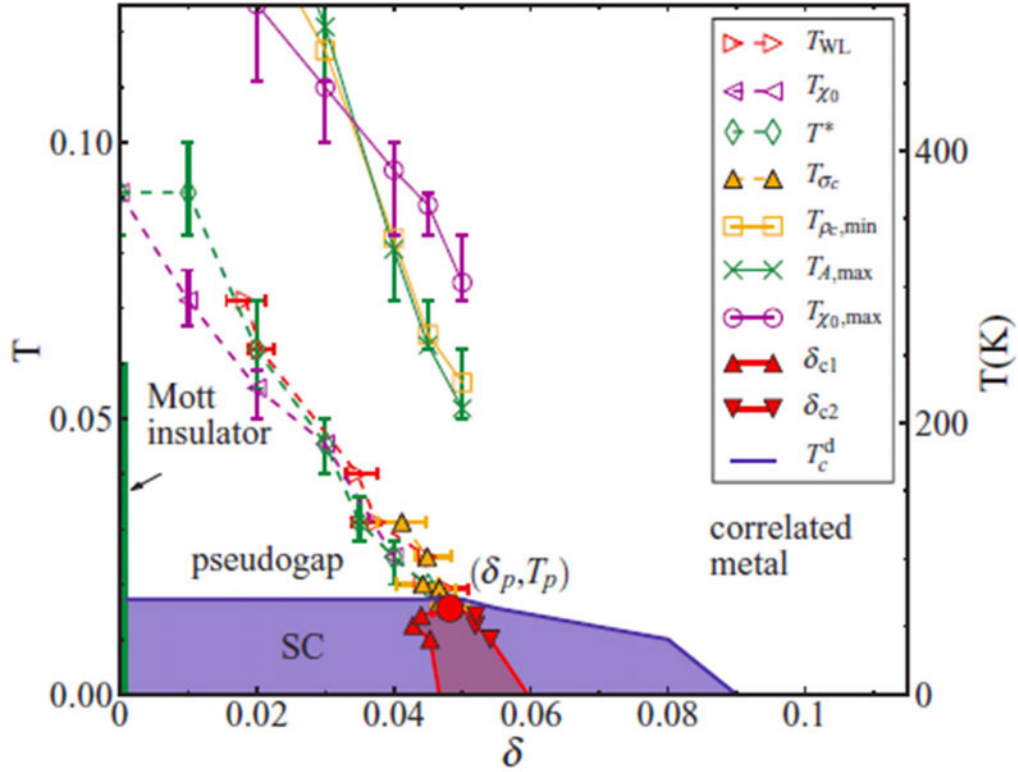


Fig. 6: Temperature vs. doping phase diagram for the nearest-neighbor Hubbard model for $U = 6.2$. The vertical axis on the left-hand side is in units of hopping. The light blue region delineates the superconducting phase. In the absence of that phase, i.e., in the metastable normal state, one finds a first-order coexistence region between the two red lines terminating at the critical point (δ_p, T_p) . The pseudogap phase is near half-filling. Red right-pointing arrows: maximum of the charge compressibility (Widom line); purple left-pointing arrows: inflection point of the spin susceptibility; green diamonds (T^*): inflection point in the zero-frequency local density of states; orange triangles: inflection point in $\sigma_c(\mu)$; orange squares: minimum in the c -axis resistivity; green crosses: maximum in the zero frequency density of states; purple circles: maximum of the spin susceptibility. Figure taken from Ref. [223].

because at larger doping the maximum simply disappears. One recovers a Pauli-like susceptibility at these dopings.³ The purple dashed line with left-pointing arrows identifies, at lower temperature, the inflection point of the susceptibility. It is very close to other crossover lines that all originate at the critical point (δ_p, T_p) of the first-order transition. The appearance of many crossover lines that all merge to terminate at the critical point is a very general phenomenon in phase transitions. It has been recently called a Widom line in the context of supercritical fluids [225] and G. Sordi has suggested that the same concept applies for the electron fluid [219]. In the latter case, a thermodynamic quantity, namely the compressibility, has a maximum at the right-pointing arrows along the red dashed line and dynamical quantities, such as the density of states, the conductivity and so on, have rapid crossovers, all analogous to the supercritical fluid

temperature. It differs from T^* in Fig. 4.

³The lines at high temperature really end before zero temperature. They should not be extrapolated to zero temperature as is often done in other theoretical work.

case.

It is noteworthy that above the crossover line where a c -axis resistivity minimum occurs (orange squares in Fig. 6), the temperature dependence is almost linear. In that regime, the c -axis resistivity can exceed the appropriate version of the Mott-Ioffe-Regel criterion [226]. Also, at zero-temperature in the pseudogap phase, it was demonstrated that very small orthorhombicity leads to very large conductivity anisotropy in a doped Mott insulator. This called electronic dynamical nematicity [227]. Such a very large anisotropy is observed experimentally in YBCO [228].

From the microscopic point of view, the probability that the plaquette is in a singlet state with four electrons increases rapidly as temperature decreases, reaching values larger than 0.5 at the lowest temperatures. The inflection point as a function of temperature of the probability for the plaquette singlet coincides with the Widom line [219].

From the point of view of this analysis, the pseudogap is a distinct phase, separated from the correlated metal by a first-order transition. It is an unstable phase at low temperature since it appears only if we suppress antiferromagnetism and superconductivity. Nevertheless, as in the case of the Mott transition, the crossovers at high-temperatures are remnants of the first-order transition. The phase transition is in the same universality class as the liquid-gas transition. As U increases, the critical point moves to larger doping and lower temperature. These calculations were for values of U very close to the Mott transition so that one could reach temperatures low enough that the first-order transition is visible. Although one sees crossover phenomena up to very large values of U , the possibility that the first-order transition turns into a quantum critical point cannot be rejected.

In summary for this section, one can infer from the plaquette studies that even with antiferromagnetic correlations that can extend at most to the first-neighbors, one can find a pseudogap and, as I discuss below, d -wave superconductivity. The pseudogap mechanism in this case is clearly related to short-range Mott physics, not to AFM correlation lengths that exceed the thermal de Broglie wavelength. Similarly, the pairing comes from the exchange interaction J , but that does not necessarily mean long wavelength antiferromagnetic correlations.

5.3 Superconducting state

When the interaction U is larger than that necessary to lead to a Mott insulator at half-filling, d -wave superconductivity has many features that are very non-BCS like. That is the topic of this section.

But first, is there d -wave superconductivity in the one-band Hubbard model or its strong-coupling version, the t - J model? Many-methods suggest that there is [114, 128, 230–232]. But there is no unanimity [233]. For reviews, see for example Refs. [74, 108, 115]. In DCA with large clusters and finite-size study, Jarrell's group has found convincing evidence of d -wave superconductivity [145] at $U = 4t$. This is too small to lead to a Mott insulator at half-filling but even for U below the Mott transition there was still sometimes some dispute regarding the existence of superconductivity. For larger U and 8-site clusters [229] one finds d -wave superconductivity and a pseudogap. It is remarkable that very similar results were obtained earlier

with the variational cluster approximation on various size clusters [234–236] and with CDMFT on 2×2 plaquettes [172, 216, 224]. With a realistic band-structure, the competition between superconductivity and antiferromagnetism can be studied and the asymmetry between the hole and electron-doped cuprates comes out clearly [172, 234–236].

Let us move back to finite temperature results. Fig. 6 shows that the superconducting phase appears in a region that is not delimited by a dome, as in experiment, but that instead the transition temperature T_c^d becomes independent of doping in the pseudogap region [224]. There is a first-order transition to the Mott insulator at half-filling. One expects that a mean-field treatment overestimates the value of T_c^d . That transition temperature T_c^d in the pseudogap region is interpreted as the temperature at which local pairs form. This could be the temperature at which a superconducting gap appears in tunneling experiments [214, 237] without long-range phase coherence. Many other experiments suggest the formation of local pairs at high temperature in the pseudogap region [238–242]. Note, however, that T_c^d is not the same as the pseudogap temperature. *The two phenomena are distinct* [224].

It is important to realize the following non-BCS feature of strongly correlated superconductivity. The saturation of T_c^d at low temperature occurs despite the fact that the order parameter has a dome shape, vanishing as we approach half-filling [172, 224]. The order parameter is discussed further below. For now, we can ask what is the effect of the size of the cluster on T_c^d . Fig. 7(a) for an 8 site cluster [229] shows that T_c^d at half-filling is roughly 30% smaller than at optimal doping, despite the fact that the low temperature superfluid density vanishes at half-filling, as seen in Fig. 7(b). Again, this is not expected from BCS. Since it seems that extremely large clusters would be necessary to observe a dome shape with vanishing T_c^d at infinitesimal doping, it would be natural to conclude that long wavelength superconducting or antiferromagnetic fluctuations are necessary to reproduce the experiment. The long-wavelength fluctuations that could be the cause of the decrease of T_c^d could be quantum and classical phase fluctuations [243–246], fluctuations in the magnitude of the order parameter [247], or of some competing order, such as antiferromagnetism or a charge-density wave. Evidently, the establishment of long-range order of a competing phase would also be effective [248]. Finally, in the real system, disorder can play a role [249, 250].

Another non-BCS feature of strongly correlated superconductivity appears in the single-particle density of states [224]. Whereas in BCS the density of states is symmetrical near zero frequency, Fig. 7(c) demonstrates that the strong asymmetry present in the pseudogap normal state (dashed red line) survives in the superconducting state. The asymmetry is clearly a property of the Mott insulator since it is easier to remove an electron ($\omega < 0$) than to add one ($\omega > 0$). Very near $\omega = 0$, all the densities of states in Fig. 7(c) are qualitatively similar since they are dictated by the symmetry-imposed nodes that are an emergent property of d -wave superconductors. Once the normal state is a correlated metal, for example at doping $\delta = 0.06$, the (particle-hole) symmetry is recovered.

The correlated metal leads then to superconducting properties akin to those of spin-fluctuation mediated BCS superconductivity. For example, in the overdoped regime superconductivity disappears concomitantly with the low frequency peak of the local spin susceptibility [149]. But in

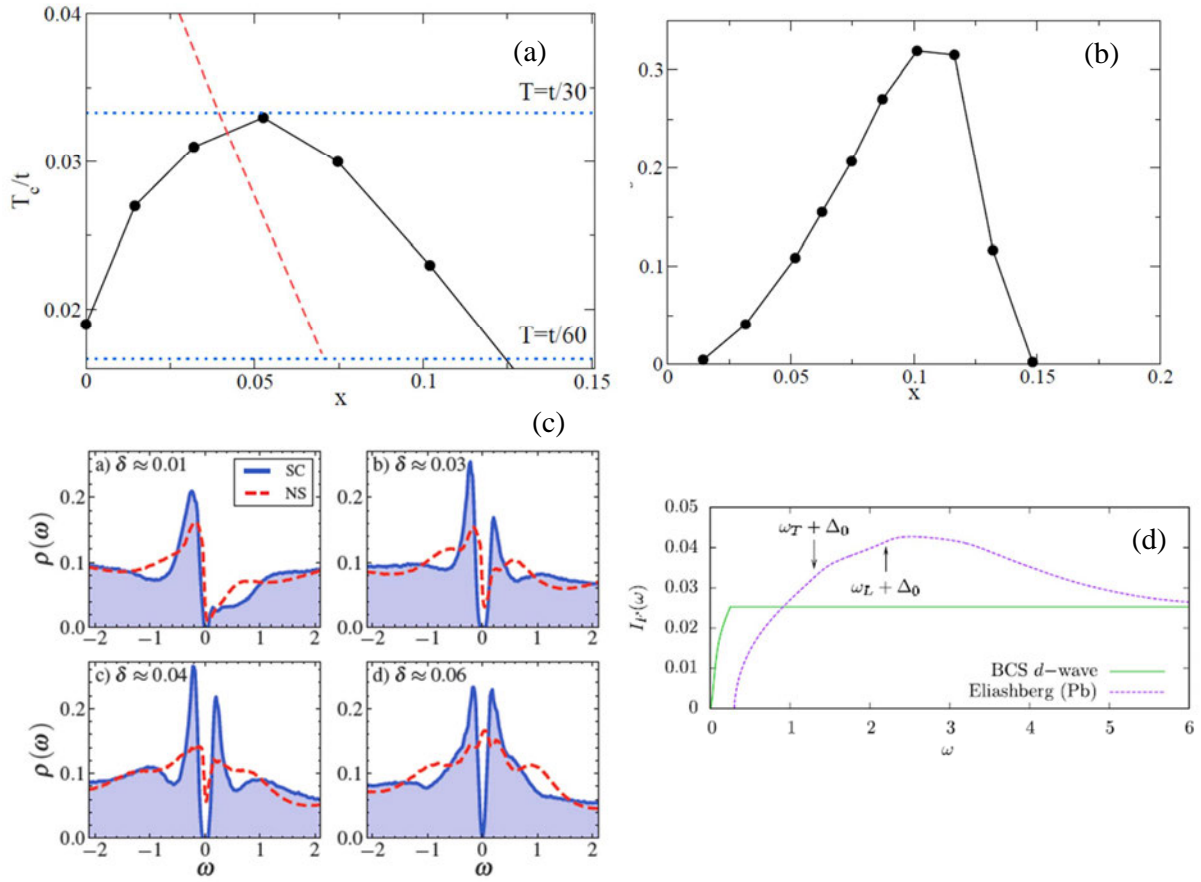


Fig. 7: (a) Superconducting critical temperature of the Hubbard model with nearest neighbor hopping calculated for $U = 6t$ using the 8-site DCA [229]. Dashed red line denotes a crossover to the normal state pseudogap. Dotted blue lines indicate the range of temperatures studied. Note that the temperature axis does not begin at zero. Figure from Ref. [229]. (b) Superfluid stiffness at $T = t/60$. Figure from Ref. [229]. (c) Low frequency part of the local density of states $\rho(\omega)$ at $U = 6.2t$, $T = 1/100$ for the normal state and the superconducting state (red dashed and blue solid lines). Figure from Ref. [224]. (d) Cumulative order parameter, i.e., the integral of the anomalous Green function (or Gork'ov function) $I_F(\omega)$. The dashed green line is $I_F(\omega)$ for a d -wave BCS superconductor with a cutoff at $\omega_c = 0.5$, which plays the role of the Debye frequency. In that case, the indices i and j in Eq. (23) are near-neighbor. The magenta line is extracted [149] from Eliashberg theory for Pb in Ref. [89]. Frequencies in that case are measured in units of the transverse phonon frequency, ω_T . The scale of the vertical axis is arbitrary. For the s -wave superconductors, one takes $i = j$ in Eq. (23). Figure from Ref. [149].

the underdoped regime, where there is a pseudogap, the difference between the pairing mechanism in a doped Mott insulator and the pairing mechanism in a doped fluctuating itinerant antiferromagnet comes out very clearly when one takes into account nearest-neighbor repulsion [90]. Indeed, the doped Mott insulator is much more resilient to near-neighbor repulsion than a spin-fluctuation mediated BCS superconductor, for reasons that go deep into the nature of superconductivity in a doped Mott insulator. This is an important result that goes much beyond the mean-field arguments of Eqs. (2) to (5). In this approach, when there is near-neighbor repulsion, one finds that superconductivity should disappear when $V > J$. In cuprates, taking

the value of the near-neighbor Coulomb interaction with a relative dielectric constant of order 10 we estimate that V , the value of near-neighbor repulsion, is of order $V \approx 400$ meV while $J \approx 130$ meV. So, from the mean-field point of view, superconductivity would not occur in the hole-doped cuprates under such circumstances.

To understand the resilience of strongly correlated superconductivity to near-neighbor repulsion V , we need to worry about the dynamics of pairing. To this end, consider the function $I_F(\omega)$, defined through the integral

$$I_F(\omega) = - \int_0^\omega \frac{d\omega'}{\pi} \text{Im } F_{ij}^R(\omega'), \quad (23)$$

where F^R is the retarded Gork'ov function (or off-diagonal Green function in the Nambu formalism) defined in imaginary time by $F_{ij} \equiv -\langle T c_{i\uparrow}(\tau) c_{j\downarrow}(0) \rangle$ with i and j nearest-neighbors. The infinite frequency limit of $I_F(\omega)$ is equal to $\langle c_{i\uparrow} c_{j\downarrow} \rangle$ which in turn is proportional to the d -wave order parameter ψ . As should become clear below, $I_F(\omega)$ is useful to estimate the frequencies that are relevant for pair binding. The name ‘‘cumulative order parameter’’ for $I_F(\omega)$ [90] is suggestive of the physical content of that function.

Fig. 7(d) illustrates the behavior of $I_F(\omega)$ in well known cases. The dashed green line is $I_F(\omega)$ for a d -wave BCS superconductor with a cutoff at $\omega_c = 0.5$. In BCS theory, that would be the Debye frequency. In BCS then, $I_F(\omega)$ is a monotonically increasing function of ω that reaches its asymptotic value at the BCS cutoff frequency ω_c [149]. The magenta line in Fig. 7(d) is obtained from Eliashberg theory for Pb in Ref. [89]. The two glitches before the maximum correspond to the transverse, ω_T , and longitudinal, ω_L , peaks in the phonon density of states. In the Eliashberg approach, which includes a retarded phonon interaction as well as the Coulomb pseudopotential μ^* that represents the repulsive electron-electron interaction [149], the function overshoots its asymptotic value at frequencies near the main phonon frequencies before decaying to its final value. The decrease occurs essentially because of the presence of the repulsive Coulomb pseudopotential, as one can deduce [90] from the examples treated in Ref. [89].

The resilience of strongly correlated superconductors to near-neighbor repulsion is best illustrated by Figs. 8(a) to (c). Each panel illustrates the order parameter as a function of doping. The dome shape that we alluded to earlier appears in panels (b) and (c) for U larger than the critical value necessary to obtain a Mott insulator at half-filling. At weak coupling, $U = 4t$ in panel (a), the dome shape appears if we allow antiferromagnetism to occur [172]. Panel (a) illustrates the sensitivity of superconductivity to near-neighbor repulsion V at weak coupling. At $V/U = 1.5/4$ superconductivity has disappeared. By contrast, at strong coupling, $U = 8t$ in panel (b), one notices that for V twice as large as in the previous case and for the same ratio $V/U = 3/8$ superconductivity is still very strong. In fact, the order parameter is not very sensitive to V in the underdoped regime. Sensitivity to V occurs mostly in the overdoped regime, which is more BCS-like. The same phenomena are observed in panel (c) for $U = 16t$.

To understand the behavior of the order parameter in the strongly correlated case, let us return to the cumulative order parameter. If we define the characteristic frequency ω_F as the frequency at which the cumulative order parameter reaches half of its asymptotic value, one can check that

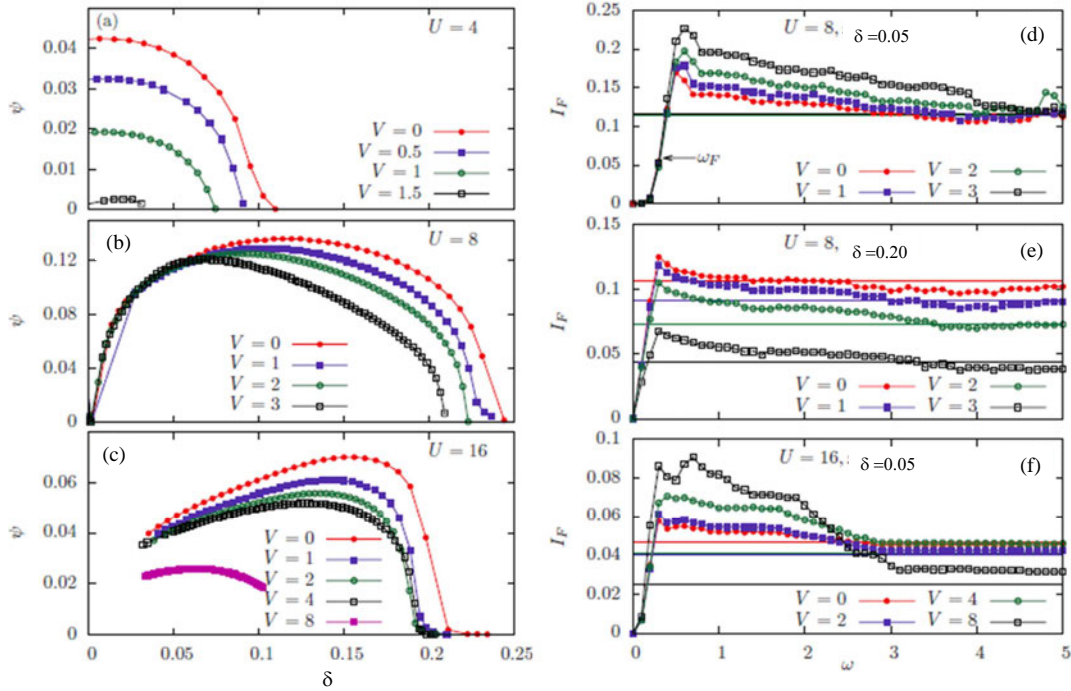


Fig. 8: All results for these figures were obtained with CDMFT and the exact diagonalization solver with the bath parametrization defined in Ref. [90]. The three panels on the left are for the d -wave order parameter ψ obtained from the off-diagonal component of the lattice Green function as a function of cluster doping for (a) $U = 4t$, (b) $U = 8t$ and (c) $U = 16t$ and various values of V . The three panels on the right represent the integral of the anomalous Green function (or Gork'ov function) $I_F(\omega)$ obtained after extrapolation to $\eta = 0$ of $\omega + i\eta$ for several values of V at (d) $U = 8t$, $\delta = 0.05$ (e) $U = 8t$, $\delta = 0.2$ (f) $U = 16t$, $\delta = 0.05$ with δ the value of doping. Frequency is measured in energy units with $t = 1$. The asymptotic value of the integral, $I_F(\infty)$, equal to the order parameter, is shown as horizontal lines. $I_F(\omega)$ is the cumulative order parameter defined by Eq. (23). The characteristic frequency ω_F is defined as the frequency at which $I_F(\omega)$ is equal to half of its asymptotic value. The horizontal arrow in panel (d) indicates how ω_F is obtained.

this frequency scales as J . Hence, not only does the optimal value of the order parameter scale as J , but so does the characteristic frequency over which the order parameter builds up.

The qualitative aspects of the effects of V , are clearer if we focus on the largest value $U = 16t$ in panel (f) for the underdoped case $\delta = 0.05$. The red curve with filled circles is for $V = 0$. For $V = 8t$, the black curve with open squares shows that the maximum of the cumulative order parameter is larger by roughly a factor of 2 than for the $V = 0$ case. This is because at strong coupling V also contributes to the effective J and thus to attraction at low frequencies. Indeed, in the presence of V , J increases from $4t^2/U$ at $V = 0$ to $4t^2/(U - V)$. That increase can be understood by considering two singly-occupied neighboring spins on a square lattice. If all other sites are occupied, the contribution to the potential energy from V is $7V$. When one of the two electrons is placed on the same site as its neighbor, they together now have three neighbors, hence the potential energy contribution from V is $6V$. The energy denominator in second-order degenerate perturbation theory thus becomes $U - 7V + 6V = U - V$, which explains

the increased value of J in the presence of V . The ratio of the effective J at $U = 16t$, $V = 8t$ to that at $U = 16t$, $V = 0$ is $U/(U - V) = 2$, which explains the observed increase in the maximum value of $I_F(\omega)$ by a factor of two. At larger frequencies, however, V is detrimental to superconductivity, leading to a Coulomb pseudopotential μ^* that reduces the value of the order parameter, just as in the case of lead described above. The effect of μ^* is largest when V is largest, as expected. Overall, the low-frequency contribution of J to the increase in binding is compensated by the high-frequency detrimental effect of V .

It is also remarkable that the frequency scale over which the cumulative order parameter reaches its asymptotic value seems to equal a few times J , just as it equals a few times the largest phonon frequency in Fig. 7(d) for the Eliashberg theory of lead. This observation is consistent with the fact that the order parameter reaches its asymptotic value for $U = 16t$ in Fig 8(f) at a frequency roughly half that where the asymptotic regime is reached for $U = 8t$ in Fig 8(d). Indeed, at $V = 0$, J is smaller by a factor of two when U increases by a factor of two.

By comparing Fig. 7(d) for the BCS and Eliashberg cases with Fig. 8(e) for the overdoped case $\delta = 0.20$ and Fig. 8(c) for the underdoped case $\delta = 0.05$, both for $U = 8t$, one verifies that the overdoped case is more BCS-like. This is consistent with the greater sensitivity of the order parameter to V that one can observe in the overdoped regime of Figs. 8(b) and (c).

Let us end with one of the most striking properties of strongly-correlated superconductivity. Layered organic superconductors of the κ -BEDT family can be modelled by the one-band Hubbard model on an anisotropic triangular lattice at half-filling [13, 251]. By changing pressure, one can tune the normal state through a Mott transition. The metallic state is at high pressure. At low temperature, pressure changes the insulating state, which can be either antiferromagnetic [11] or spin liquid [252], to a superconducting state with a non s -wave order parameter. See Ref. [253] for a review. One finds experimentally that the superconducting T_c is largest at the transition, in other words, closest to the insulating phase. In addition, T_c is larger in the compounds that have the largest mass renormalization in the normal metallic state, i.e., in the most strongly correlated ones [13]. All of this is highly counter-intuitive. I have run out of space to explain the theoretical situation on this issue. Let us just remark that at $T = 0$ with an exact diagonalization solver, one finds with quantum cluster methods [254] that indeed the order-parameter is largest near the first-order Mott transition [255]. If T_c scales with the order parameter, this type of approach thus reproduces a counter-intuitive result. The existence of unconventional superconductivity in the Hubbard model on the anisotropic triangular lattice is, however, disputed [256].

In the model for the cuprates, the value of the order parameter at optimal doping increases with U starting from small U and reaches a maximum at intermediate coupling before decreasing with J at large U . In other words, as a function of U , the largest value that the order parameter can take is for $U \approx 6t$ where the Mott transition occurs at half-filling. Analogously, as a function of doping, at $U = 6.2$, the maximum T_c occurs near the critical doping for the pseudogap to correlated metal transition [224].

6 Conclusion

Progress with numerical methods, especially cluster generalizations of DMFT, has shown in recent years that much of the physics of strongly correlated superconductors is contained in the one-band Hubbard model. Confidence in the method comes from extensive benchmarking, and from the agreement at intermediate coupling with TPSC, which is also benchmarked and valid up to intermediate coupling. The fact that both the cuprates and the layered organics are well described gives additional confidence in the validity of the approach. Much physical insight can be gained by these methods. In the future, it will be useful to use them to discriminate between the various versions of mean-field theories based on slave-particle approaches. A mean-field approach that would contain most of the features of numerical approaches might help us to gain further insight into the problem. Variational wave functions, even if treated with numerical methods, are also helpful to this end.

Much remains to be done. In the one-band model, it is necessary to develop methods that allow us to reliably investigate long-wavelength instabilities, such as the CDW observed for the cuprates. Investigating theoretically all the details of the highly unusual superconducting properties of the layered organics also remains to be done. And the description of all features of the cuprates at lower temperature calls for even more extensive studies of the three-band model [115, 257–259].

Acknowledgments

I am especially grateful to David Sénéchal, Giovanni Sordi, Patrick Sémon, Dominic Bergeron, Yury Vilk, Bumsoo Kyung, Kristjan Haule and to many other collaborators on some of the work discussed above, including V. Bouliane, M. Capone, A. Day, M. Civelli, B. Davoudi, S. Kancharla, G. Kotliar, J.-S. Landry, P.-L. Lavertu, M.-A. Marois, S. Roy. This work was partially supported by the Tier I Canada Research Chair Program (A.-M.S.T.) by CFI, MELS, Calcul Québec and Compute Canada.

References

- [1] J.H. de Boer and E.J.W. Verwey, *Proceedings of the Physical Society* **49**, 59 (1937)
- [2] M. Imada, A. Fujimori, and Y. Tokura, *Rev. Mod. Phys.* **70**, 1039 (1998)
- [3] F. Steglich, J. Aarts, C.D. Bredl, W. Lieke, D. Meschede, W. Franz, and H. Schäfer, *Physical Review Letters* **43**, 1892 (1979)
- [4] D. Jérôme, A. Mazaud, M. Ribault, and K. Bechgaard, *J. Phys. Lett.* **41**, L95 (1980)
- [5] J. Bednorz and K. Müller, *Z. Phys. B* **64**, 189 (1986)
- [6] J. Williams, A.J. Schultz, U. Geiser, K.D. Carlson, A.M. Kini, H. Wang, W.K. Kwok, M.-H. Whangbo, and J.E. Schriber, *Science* **252**, 1501 (1991)
- [7] D. Jérôme, *Science* **252**, 1509 (1991)
- [8] Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, *J. Am. Chem. Soc.* **130**, 3296 (2008)
- [9] H. Ito, T. Ishiguro, M. Kubota, and G. Saito, *J. Phys. Soc. Japan* **65**, 2987 (1996)
- [10] K. Kanoda, *Physica C: Superconductivity* **282-287**, part 1, 299 (1997)
- [11] S. Lefebvre, P. Wzietek, S. Brown, C. Bourbonnais, D. Jérôme, C. Mézière, M. Fourmigué, and P. Batail, *Phys. Rev. Lett.* **85**, 5420 (2000)
- [12] P. Limelette, P. Wzietek, S. Florens, A. Georges, T.A. Costi, C. Pasquier, D. Jérôme, C. Mézière, and P. Batail, *Phys. Rev. Lett.* **91**, 016401 (2003)
- [13] B.J. Powell and R.H. McKenzie, *J. Phys. Cond. Mat.* **18**, R827 (2006)
- [14] J. Hubbard, *Proc. Roy. Soc. London A* **276**, 238 (1963)
- [15] J. Kanamori, *Prog. Theor. Phys.* **30**, 275 (1963)
- [16] M.C. Gutzwiller, *Phys. Rev. Lett.* **10**, 159 (1963)
- [17] P. Nozières and J.M. Luttinger, *Phys. Rev.* **127**, 1423 (1962)
- [18] J.M. Luttinger and P. Nozières, *Phys. Rev.* **127**, 1431 (1962)
- [19] A.B. Harris and R.V. Lange, *Phys. Rev.* **157**, 295 (1967)
- [20] W. Metzner, *Phys. Rev. B* **43**, 8549 (1991)
- [21] M. Bartkowiak and K.A. Chao, *Phys. Rev. B* **46**, 9228 (1992)
- [22] S. Pairault, D. Sénéchal, and A.-M.S. Tremblay, *Phys. Rev. Lett.* **80**, 5389 (1998)

- [23] S. Pairault, D. Sénéchal, and A.-M.S. Tremblay, Eur. Phys. J. B **16**, 85 (2000)
- [24] C. Gros, R. Joynt, and T.M. Rice, Phys. Rev. B **36**, 381 (1987)
- [25] P.W. Anderson, Phys. Rev. **115**, 2 (1959)
- [26] P.W. Anderson, Science **235**, 1196 (1987)
- [27] M.B.J. Meinders, H. Eskes, and G.A. Sawatzky, Phys. Rev. B **48**, 3916 (1993)
- [28] P.W. Anderson, Science **177**, 393 (1972)
- [29] R.B. Laughlin: *A Different Universe: Reinventing Physics from the Bottom Down* (Basic Books, 2006)
- [30] D. Pines: *Emergent Behavior in Quantum Matter* (2010)
<http://www.learner.org/courses/physics/unit/text.html?unit=8&secNum=0>
- [31] A. Reymbaut, unpublished (2013)
- [32] N.D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1133 (1966)
- [33] P.C. Hohenberg, Phys. Rev. **158**, 383 (1967)
- [34] G. Toulouse, Commun. Phys. (1977)
- [35] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [36] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992)
- [37] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [38] J.R. Schrieffer, X.G. Wen, and S.C. Zhang: Phys. Rev. B **39**, 11663 (1989)
- [39] J.D. Reger and A.P. Young, Phys. Rev. B **37**, 5978 (1988)
- [40] D. Vaknin, S.K. Sinha, D.E. Moncton, D.C. Johnston, J.M. Newsam, C.R. Safinya, and H.E. King, Phys. Rev. Lett. **58**, 2802 (1987)
- [41] M. Takahashi, J. Phys. C: Solid State Physics **10**, 1289 (1977)
- [42] A.H. MacDonald, S.M. Girvin, and D. Yoshioka, Phys. Rev. B **37**, 9753 (1988)
- [43] J.-Y.P. Delannoy, M.J.P. Gingras, P.C.W. Holdsworth, and A.-M.S. Tremblay, Phys. Rev. B **72**, 115114 (2005)
- [44] H.-Y. Yang, A.M. Läuchli, F. Mila, and K.P. Schmidt, Phys. Rev. Lett. **105**, 267204 (2010)

- [45] B. Dalla Piazza, M. Mourigal, M. Guarise, H. Berger, T. Schmitt, K.J. Zhou, M. Grioni, and H.M. Rønnow, *Phys. Rev. B* **85**, 100508 (2012)
- [46] T. Holstein and H. Primakoff, *Phys. Rev.* **58**, 1098 (1940)
- [47] A.V. Chubukov and S. Sachdev, *Phys. Rev. Lett.* **71**, 169 (1993)
- [48] F.D.M. Haldane, *Phys. Rev. Lett.* **50**, 1153 (1983)
- [49] A.W. Sandvik: *Lecture notes for course given at the 14th Training Course in Physics of Strongly Correlated Systems, Salerno* (Vietri sul Mare, Italy, October 2009)
<http://arxiv.org/abs/1101.3281>
- [50] M. Campostrini, M. Hasenbusch, A. Pelissetto, P. Rossi, and E. Vicari, *Phys. Rev. B* **65**, 144520 (2002)
- [51] E. Kozik, E. Burovski, V.W. Scarola, and M. Troyer, *Phys. Rev. B* **87**, 205102 (2013)
- [52] N. Prokof'ev, B. Svistunov, and I. Tupitsyn, *Phys. Lett. A* **238**, 253 (1998)
- [53] M. Boninsegni, N.V. Prokof'ev, and B.V. Svistunov, *Phys. Rev. E* **74**, 036701 (2006)
- [54] S. Sorella, *Phys. Rev. B* **71**, 241103 (2005)
- [55] T. Giamarchi and C. Lhuillier, *Phys. Rev. B* **43**, 12943 (1991)
- [56] E. Manousakis, *Rev. Mod. Phys.* **63**, 1 (1991)
- [57] <http://pitp.physics.ubc.ca/confs/sherbrooke/archives.html>
- [58] <http://pitp.physics.ubc.ca/confs/sherbrooke2012/archives.html>
- [59] S. Chakravarty, B.I. Halperin, and D.R. Nelson, *Phys. Rev. B* **39**, 2344 (1989)
- [60] B. Keimer, N. Belk, R. J. Birgeneau, A. Cassanho, C.Y. Chen, M. Greven, M.A. Kastner, A. Aharony, Y. Endoh, R.W. Erwin, and G. Shirane, *Phys. Rev. B* **46**, 14034 (1992)
- [61] P. Monthoux and D. Pines, *Phys. Rev. B* **47**, 6069 (1993)
- [62] T. Moriya: *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, Heidelberg, 1985)
- [63] N.E. Bickers, D.J. Scalapino, and S.R. White, *Phys. Rev. Lett.* **62**, 961 (1989)
- [64] N.E. Bickers and D.J. Scalapino, *Ann. Phys. (USA)* **193**, 206 (1989)
- [65] D. Zanchi and H.J. Schulz, *Phys. Rev. B* **61**, 13609 (2000)

- [66] C. Honerkamp, M. Salmhofer, N. Furukawa, and T.M. Rice, Phys. Rev. B **63**, 035109 (2001)
- [67] R. Shankar, Rev. Mod. Phys. **66**, 129 (1994)
- [68] R.K. Kaul, M.A. Metlitski, S. Sachdev, and C. Xu, Phys. Rev. B **78**, 045110 (2008)
- [69] S. Sachdev, M.A. Metlitski, Y. Qi, and C. Xu, Phys. Rev. B **80**, 155129 (2009)
- [70] Y.M. Vilk and A.-M.S. Tremblay, J. Phys I (France) **7**, 1309 (1997)
- [71] S. Allen, A.-M.S. Tremblay, and Y.M. Vilk, in D. Sénéchal, C. Bourbonnais, and A.-M.S. Tremblay (eds.): *Theoretical Methods for Strongly Correlated Electrons* (2003)
- [72] A. M. S. Tremblay: *Two-Particle-Self-Consistent Approach for the Hubbard Model* in F. Mancini and A. Avella (eds.): *Strongly Correlated Systems* (Springer, 2011)
- [73] N. Bulut, D.J. Scalapino, and S.R. White, Phys. Rev. Lett. **73**, 748 (1994)
- [74] A. M.S. Tremblay, B. Kyung, and D. Sénéchal, Low Temp. Phys. **32**, 424 (2006)
- [75] K.A. Brueckner, T. Soda, P.W. Anderson, and P. Morel, Phys. Rev. **118**, 1442 (1960)
- [76] Y. M. Vilk and A.-M.S. Tremblay, J. Phys. Chem. Solids (UK) **56**, 1769 (1995)
- [77] Y. M. Vilk and A.-M.S. Tremblay, Europhys. Lett. **33**, 159 (1996)
- [78] S. Moukouri, S. Allen, F. Lemay, B. Kyung, D. Poulin, Y.M. Vilk, and A.-M.S. Tremblay, Phys. Rev. B **61**, 7887 (2000)
- [79] E.M. Motoyama, G. Yu, I.M. Vishik, O.P. Vajk, P.K. Mang, and M. Greven, Nature **445**, 186 (2007)
- [80] B. Kyung, V. Hankevych, A.-M. Daré, and A.-M.S. Tremblay, Phys. Rev. Lett. **93**, 147004 (2004)
- [81] D. Bergeron, D. Chowdhury, M. Punk, S. Sachdev, and A.-M.S. Tremblay, Phys. Rev. B **86**, 155123 (2012)
- [82] C.M. Varma, J. Low Temp. Phys. **126**, 901 (2002)
- [83] P.A. Lee, Phys. Rev. Lett. **71**, 1887 (1993)
- [84] L. Taillefer, B. Lussier, R. Gagnon, K. Behnia, and H. Aubin, Phys. Rev. Lett. **79**, 483 (1997)
- [85] G.M. Eliashberg, Soviet Phys. JETP **11**, 696 (1960)
- [86] J.P. Carbotte, Rev. Mod. Phys. **62**, 1027 (1990)

- [87] F. Marsiglio and J. Carbotte, arXiv:cond-mat/0106143 (2001)
- [88] J.R. Schrieffer: *Theory of Superconductivity* (W.A. Benjamin, New York, 1964)
- [89] D.J. Scalapino, J.R. Schrieffer, and J.W. Wilkins, Phys. Rev. **148**, 263 (1966)
- [90] D. Sénéchal, A.G.R. Day, V. Bouliane, and A.-M.S. Tremblay, Phys. Rev. B **87**, 075123 (2013)
- [91] N.E. Hussey, K. Takenaka, and H. Takagi, Phil. Mag. **84**, 2847 (2004)
- [92] M. Gurvitch and A.T. Fiory, Phys. Rev. Lett. **59**, 1337 (1987)
- [93] O. Gunnarsson, M. Calandra, and J.E. Han, Rev. Mod. Phys. **75**, 1085 (2003)
- [94] D. Bergeron, V. Hankevych, B. Kyung, and A.-M.S. Tremblay, Phys. Rev. B **84**, 085128 (2011)
- [95] W. Xu, K. Haule, and G. Kotliar: arXiv:1304.7486 (2013)
- [96] P.A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006)
- [97] S.D. Obertelli, J.R. Cooper, and J.L. Tallon, Phys. Rev. B **46**, 14928 (1992)
- [98] T. Honma and P. H. Hor, Phys. Rev. B **77**, 184520 (2008)
- [99] Y. Ando, Y. Kurita, S. Komiya, S. Ono, and K. Segawa, Phys. Rev. Lett. **92**, 197001 (2004)
- [100] D.C. Peets, D.G. Hawthorn, K.M. Shen, Y.-J. Kim, D.S. Ellis, H. Zhang, S. Komiya, Y. Ando, G.A. Sawatzky, R. Liang, D. A. Bonn, and W.N. Hardy, Phys. Rev. Lett. **103**, 087402 (2009)
- [101] P. Phillips and M. Jarrell, Phys. Rev. Lett. **105**, 199701 (2010)
- [102] C.T. Chen, F. Sette, Y. Ma, M.S. Hybertsen, E.B. Stechel, W.M.C. Foulkes, M. Schuler, S.-W. Cheong, A.S. Cooper, L.W. Rupp, B. Batlogg, Y.L. Soo, Z.H. Ming, A. Krol, and Y.H. Kao, Phys. Rev. Lett. **66**, 104 (1991)
- [103] K. Miyake, S. Schmitt-Rink, and C.M. Varma, Phys. Rev. B **34**, 6554 (1986)
- [104] G. Kotliar and J. Liu, Phys. Rev. Lett. **61**, 1784 (1988)
- [105] P.W. Anderson, P.A. Lee, M. Randeria, T.M. Rice, N. Trivedi, and F.C. Zhang, J. Phys.: Condensed Matter **16**, R755 (2004)
- [106] W. Kohn and J.M. Luttinger, Phys. Rev. Letters **15**, 524 (1965)
- [107] J. Mráz and R. Hlubina, Phys. Rev. B **67**, 174518 (2003)

- [108] S. Maiti and A.V. Chubukov: arXiv:1305.4609 (2013)
- [109] S. Raghu, S.A. Kivelson, and D.J. Scalapino, Phys. Rev. B **81**, 224505 (2010)
- [110] M.T. Beal-Monod, C. Bourbonnais, and V.J. Emery, Phys. Rev. B **34**, 7716 (1986)
- [111] D.J. Scalapino, J. Loh, E., and J.E. Hirsch, Phys. Rev. B **34**, 8190 (1986)
- [112] M.R. Norman, M. Randeria, B. Janko, and J.C. Campuzano, Phys. Rev. B **61**, 14742 (2000)
- [113] T. Moriya and K. Ueda, Rep. Prog. Phys. **66**, 1299 (2003)
- [114] P. Monthoux, D. Pines, and G.G. Lonzarich, Nature **450**, 1177 (2007)
- [115] D.J. Scalapino: Physica C **470**, S1 (2010)
- [116] S.R. Hassan, B. Davoudi, B. Kyung, and A.-M.S. Tremblay, Phys. Rev. B **77**, 094501 (2008)
- [117] D.M. Broun, Nat Phys **4**, 170 (2008)
- [118] S. Sachdev, physica status solidi (b) **247**, 537 (2010)
- [119] N. Doiron-Leyraud and L. Taillefer, Physica C **481**, 161 (2012)
- [120] L.G. Caron and C. Bourbonnais Physica B & C **143**, 453 (1986)
- [121] C. Bourbonnais and A. Sedeki, Phys. Rev. B. in press (2012)
- [122] C. Bourbonnais, Mol. Cryst. Liq. Cryst. **119**, 11 (1985)
- [123] C. Bourbonnais, in C.B. Dou and J. Zinn-Justin (eds.) *Les Houches 1991, Session LVI: Strongly interacting fermions and high T_c superconductivity* (Elsevier, 1995), pp. 307
- [124] C.J. Halboth and W. Metzner, Phys. Rev. Lett. **85**, 5162 (2000)
- [125] M. Salmhofer, C. Honerkamp, W. Metzner, and O. Lauscher, Prog. Theor. Phys. **112**, 943 (2004)
- [126] W. Metzner, M. Salmhofer, C. Honerkamp, V. Meden, and K. Schönhammer, Rev. Mod. Phys. **84**, 299 (2012)
- [127] B. Kyung, J.-S. Landry, and A.-M.S. Tremblay, Phys. Rev. B **68**, 174502 (2003)
- [128] S. Nishiyama, K. Miyake, and C. Varma, cond-mat/1302.2160v1 (2013)
- [129] N. Doiron-Leyraud, P. Auban-Senzier, S. René de Cotret, C. Bourbonnais, D. Jérôme, K. Bechgaard, and L. Taillefer, Phys. Rev. B **80**, 214531 (2009)

- [130] Y. Zhang, E. Demler, and S. Sachdev, Phys. Rev. B **66**, 094501 (2002)
- [131] S.-X. Yang, H. Fotso, S.-Q. Su, D. Galanakis, E. Khatami, J.-H. She, J. Moreno, J. Zaanen, and M. Jarrell, Phys. Rev. Lett. **106**, 047004 (2011)
- [132] S.E. Barnes, J. Phys. F **6**, 1375 (1976)
- [133] P. Coleman, Phys. Rev. B **29**, 3035 (1984)
- [134] G. Kotliar and A.E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986)
- [135] D. Yoshida, J. Phys. Soc. Japan **58**, 32 (1989)
- [136] C. Jayaprakash, H.R. Krishnamurthy, and S. Sarker, Phys. Rev. B **40**, 2610 (1989)
- [137] S. Florens and A. Georges, Phys. Rev. B **66**, 165111 (2002)
- [138] S.R. Hassan and L. de' Medici, Phys. Rev. B **81**, 035106 (2010)
- [139] P. Phillips, Rev. Mod. Phys. **82**, 1719 (2010).
- [140] D. Boies, F. A. Jackson, and T. A. M. S., Int. J. Mod. Phys. B **9**, 1001 (1995)
- [141] K.-Y. Yang, T.M. Rice, and F.-C. Zhang, Phys. Rev. B **73**, 174501 (2006)
- [142] E. Schachinger and J.P. Carbotte, Phys. Rev. B **81**, 014519 (2010)
- [143] J.P.F. LeBlanc, J.P. Carbotte, and E.J. Nicol, Phys. Rev. B **83**, 184506 (2011)
- [144] P.Sémon and G. Sordi, unpublished (2013)
- [145] T.A. Maier, M. Jarrell, T.C. Schulthess, P.R.C. Kent, and J.B. White, Phys. Rev. Lett. **95**, 237001 (2005)
- [146] M. Inui, S. Doniach, P.J. Hirschfeld, and A.E. Ruckenstein, Phys. Rev. B **37**, 2320 (1988)
- [147] P.W. Anderson, Science **316**, 1705 (2007)
- [148] T.A. Maier, D. Poilblanc, and D.J. Scalapino, Phys. Rev. Lett. **100**, 237001 (2008)
- [149] B. Kyung, D. Sénéchal, and A.-M.S. Tremblay, Phys. Rev. B **80**, 205109 (2009)
- [150] D.J. Scalapino, Physics Reports **250**, 329 (1995)
- [151] A. Avela and F. Mancini: *The Composite Operator Method (COM)* in F. Mancini and A. Avella (eds.): *Strongly Correlated Systems* (Springer, 2011)
- [152] M.H. Hettler, A.N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H.R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998)

- [153] G. Kotliar, S. Y. Savrasov, G. Pálsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001)
- [154] A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **62**, R9283 (2000)
- [155] T. Maier, M. Jarrell, T. Pruschke, and M.H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005)
- [156] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)
- [157] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [158] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003)
- [159] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003)
- [160] M. Potthoff, Eur. Phys. J. B **32**, 429 (2003)
- [161] M. Potthoff: *Self-Energy-Functional Theory*
in F. Mancini and A. Avella (eds.): *Strongly Correlated Systems* (Springer, 2011)
- [162] T.D. Stanescu, M. Civelli, K. Haule, and G. Kotliar: Ann. Phys. **321**, 1682 (2006)
- [163] D. Sénéchal: In D. Sénéchal (ed.) *High Performance Computing Systems and Applications and OSCAR symposium* (2003)
- [164] T. D. Stanescu and G. Kotliar: Phys. Rev. B **70**, 205112 (2004)
- [165] S. Sakai, G. Sangiovanni, M. Civelli, Y. Motome, K. Held, and M. Imada, Phys. Rev. B **85**, 035102 (2012)
- [166] Th. Maier, M. Potthoff and D. Sénéchal, unpublished
- [167] G. Biroli and G. Kotliar, Phys. Rev. B **65**, 155112 (2002)
- [168] K. Aryanpour, T.A. Maier, and M. Jarrell, Phys. Rev. B **71**, 037101 (2005)
- [169] G. Biroli and G. Kotliar, Phys. Rev. B **71**, 037102 (2005)
- [170] B. Kyung, G. Kotliar, and A.-M.S. Tremblay, Phys. Rev. B **73**, 205106 (2006)
- [171] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
- [172] S.S. Kancharla, B. Kyung, D. Senechal, M. Civelli, M. Capone, G. Kotliar, and A.-M.S. Tremblay, Phys. Rev. B **77**, 184516 (2008)
- [173] M. Capone, M. Civelli, S.S. Kancharla, C. Castellani, and G. Kotliar, Phys. Rev. B **69**, 195105 (2004)
- [174] M. Civelli: *Investigation of strongly correlated electron system with cellular dynamical mean-field theory*. (Ph.D. thesis, Rutgers University, 2006)

- [175] B. Kyung, S.S. Kancharla, D. Sénéchal, A.-M.S. Tremblay, M. Civelli, and G. Kotliar, Phys. Rev. B **73**, 165114 (2006)
- [176] A. Liebsch and N.-H. Tong, Phys. Rev. B **80**, 165126 (2009)
- [177] A. Liebsch and H. Ishida, J. Phys.: Condensed Matter **24**, 053201 (2012)
- [178] W. Hofstetter, Phys. Rev. Lett. **85**, 1508 (2000)
- [179] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)
- [180] D. Sénéchal: *An introduction to quantum cluster methods*, arXiv:0806.2690
- [181] D. Sénéchal: *Cluster Dynamical Mean-Field Theory* in A. Avella and F. Mancini (eds.): *Strongly Correlated Systems* (Springer, 2012)
- [182] C. J. Bolech, S.S. Kancharla, and G. Kotliar, Phys. Rev. B **67**, 075110 (2003)
- [183] T.D. Stanescu and G. Kotliar: *Cluster Dynamical Mean Field Theories* (2005)
- [184] H. Alloul, T. Ohno, and P. Mendels, Phys. Rev. Lett. **63**, 1700 (1989)
- [185] C. Renner, B. Revaz, J.-Y. Genoud, K. Kadowaki, and . Fischer, Phys. Rev. Lett. **80**, 149 (1998)
- [186] A.G. Loeser, Z.-X. Shen, D.S. Dessau, D.S. Marshall, C.H. Park, P. Fournier, and A. Kapitulnik, Science **273**, 325 (1996)
- [187] H. Ding, M.R. Norman, J.C. Campuzano, M. Randeria, A.F. Bellman, T. Yokoya, T. Takahashi, T. Mochiku, and K. Kadowaki, Phys. Rev. B **54**, R9678 (1996)
- [188] A.V. Puchkov, D.N. Basov, and T. Timusk, J. Phys.: Condensed Matter **8**, 10049 (1996)
- [189] M.R. Norman, D. Pines, and C. Kallin, Adv. Phys. **54**, 715 (2005)
- [190] T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999)
- [191] D. Sénéchal and A.-M.S. Tremblay, Phys. Rev. Lett. **92**, 126401 (2004)
- [192] C. Weber, K. Haule, and G. Kotliar, Nature Physics **6**, 574 (2010)
- [193] V. Hankevych, B. Kyung, A.-M. Daré, D. Sénéchal, and A.-M.S. Tremblay, J. Phys. Chem. Sol. **67**, 189 (2006)
- [194] H. Alloul, P. Mendels, H. Casalta, J.F. Marucco, and J. Arabski, Phys. Rev. Lett. **67**, 3140 (1991)

- [195] J. Xia, E. Schemm, G. Deutscher, S.A. Kivelson, D.A. Bonn, W.N. Hardy, R. Liang, W. Siemons, G. Koster, M.M. Fejer, and A. Kapitulnik, *Phys. Rev. Lett.* **100**, 127002 (2008)
- [196] T. Wu, H. Mayaffre, S. Krämer, M. Horvatič, C. Berthier, W.N. Hardy, R. Liang, D.A. Bonn, and M.-H. Julien, *Nature* **477**, 191 (2011)
- [197] D. LeBoeuf, S. Kramer, W. N. Hardy, R. Liang, D.A. Bonn, and C. Proust, *Nat. Phys* **9**, 79 (2013)
- [198] J. Chang, E. Blackburn, A.T. Holmes, N.B. Christensen, J. Larsen, J. Mesot, R. Liang, D.A. Bonn, W. N. Hardy, A. Watenphul, M. v. Zimmermann, E.M. Forgan, and S.M. Hayden, *Nature Physics* **8**, 871 (2012)
- [199] A. Shekhter, B.J. Ramshaw, R. Liang, W.N. Hardy, D.A. Bonn, F.F. Balakirev, R.D. McDonald, J.B. Betts, S.C. Riggs, and A. Migliori, *Nature* **498**, 75 (2013)
- [200] R. Daou, J. Chang, D. LeBoeuf, O. Cyr-Choiniere, F. Laliberte, N. Doiron-Leyraud, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy, and L. Taillefer, *Nature* **463**, 519 (2010)
- [201] H. Alloul, F. Rullier-Albenque, B. Vignolle, D. Colson, and A. Forget *EPL* **91**, 37005 (2010)
- [202] H. Alloul, arXiv:1302.3473 (2013)
- [203] S.A. Kivelson, I.P. Bindloss, E. Fradkin, V. Oganessian, J.M. Tranquada, A. Kapitulnik, and C. Howald, *Rev. Mod. Phys.* **75**, 1201 (2003)
- [204] S. Chakravarty, R.B. Laughlin, D.K. Morr, and C. Nayak, *Phys. Rev. B* **63**, 094503 (2001)
- [205] J. Chang, C. Niedermayer, R. Gilardi, N.B. Christensen, H.M. Rønnow, D.F. McMorrow, M. Ay, J. Stahn, O. Sobolev, A. Hiess, S. Pailhes, C. Baines, N. Momono, M. Oda, M. Ido, and J. Mesot, *Phys. Rev. B* **78**, 104525 (2008)
- [206] G. Ghiringhelli, M. Le Tacon, M. Minola, S. Blanco-Canosa, C. Mazzoli, N.B. Brookes, G.M. De Luca, A. Frano, D.G. Hawthorn, F. He, T. Loew, M.M. Sala, D. C. Peets, M. Salluzzo, E. Schierle, R. Sutarto, G. A. Sawatzky, E. Weschke, B. Keimer, and L. Braicovich, *Science* **337**, 821 (2012)
- [207] J. Chang, R. Daou, C. Proust, D. LeBoeuf, N. Doiron-Leyraud, F. Laliberté, B. Pingault, B.J. Ramshaw, R. Liang, D.A. Bonn, W.N. Hardy, H. Takagi, A.B. Antunes, I. Sheikin, K. Behnia, and L. Taillefer, *Phys. Rev. Lett.* **104**, 057005 (2010)
- [208] N. Doiron-Leyraud, C. Proust, D. LeBoeuf, J. Levallois, J.-B. Bonnemaïson, R. Liang, D.A. Bonn, W.N. Hardy, and L. Taillefer, *Nature* **447**, 565 (2007)

- [209] M.J. Lawler, K. Fujita, J. Lee, A.R. Schmidt, Y. Kohsaka, C. K. Kim, H. Eisaki, S. Uchida, J.C. Davis, J.P. Sethna, and E.-A. Kim, *Nature* **466**, 347 (2010)
- [210] N. Doiron-Leyraud, S. Lepault, O. Cyr-Choiniere, B. Vignolle, G. Grissonnanche, F. Laliberte, J. Chang, N. Barisic, M.K. Chan, L. Ji, X. Zhao, Y. Li, M. Greven, C. Proust, and L. Taillefer, *arXiv* (2012)
- [211] S.E. Sebastian, N. Harrison, R. Liang, D.A. Bonn, W.N. Hardy, C.H. Mielke, and G. G. Lonzarich, *Phys. Rev. Lett.* **108**, 196403 (2012)
- [212] Y. Sidis and P. Bourges, *arXiv* (2013)
- [213] C.M. Varma, *Phys. Rev. B* **55**, 14554 (1997)
- [214] K.K. Gomes, A.N. Pasupathy, A. Pushp, S. Ono, Y. Ando, and A. Yazdani, *Nature* **447**, 569 (2007)
- [215] C. Huscroft, M. Jarrell, T. Maier, S. Moukouri, and A.N. Tahvildarzadeh, *Phys. Rev. Lett.* **86**, 139 (2001)
- [216] K. Haule and G. Kotliar, *Phys. Rev. B* **76**, 104509 (2007)
- [217] M. Ferrero, P.S. Cornaglia, L. De Leo, O. Parcollet, G. Kotliar, and A. Georges, *Phys. Rev. B* **80**, 064501 (2009)
- [218] E. Gull, M. Ferrero, O. Parcollet, A. Georges, and A.J. Millis, *Phys. Rev. B* **82**, 155101 (2010)
- [219] G. Sordi, P. Sémon, K. Haule, and A.-M.S. Tremblay, *Scientific Reports* **2** (2012)
- [220] P. Werner, E. Gull, O. Parcollet, and A.J. Millis, *Phys. Rev. B* **80**, 045120 (2009)
- [221] G. Sordi, K. Haule, and A.-M.S. Tremblay, *Phys. Rev. Lett.* **104**, 226402 (2010)
- [222] G. Sordi, K. Haule, and A.-M.S. Tremblay, *Phys. Rev. B* **84**, 075161 (2011)
- [223] G. Sordi, P. Sémon, K. Haule, and A.-M.S. Tremblay, *Phys. Rev. B* **87**, 041101 (2013)
- [224] G. Sordi, P. Sémon, K. Haule, and A.-M.S. Tremblay, *Phys. Rev. Lett.* **108**, 216401 (2012)
- [225] L. Xu, P. Kumar, S.V. Buldyrev, S.-H. Chen, P.H. Poole, F. Sciortino, and H.E. Stanley, *PNAS* **102**, 16558 (2005)
- [226] T. Ito, H. Takagi, S. Ishibashi, T. Ido, and S. Uchida, *Nature* **350**, 596 (1991)
- [227] S. Okamoto, D. Sénéchal, M. Civelli, and A.-M.S. Tremblay, *Phys. Rev. B* **82**, 180511 (2010)

- [228] Y. Ando, K. Segawa, S. Komiya, and A.N. Lavrov, Phys. Rev. Lett. **88**, 137005 (2002)
- [229] E. Gull, O. Parcollet, and A.J. Millis, Physical Review Letters **110**, 216405 (2013)
- [230] A. Paramekanti, M. Randeria, and N. Trivedi, Phys. Rev. B **70**, 054504 (2004)
- [231] S. Sorella, G.B. Martins, F. Becca, C. Gazza, L. Capriotti, A. Parola, and E. Dagotto, Phys. Rev. Lett. **88**, 117002 (2002)
- [232] D. Poilblanc and D.J. Scalapino, Phys. Rev. B **66**, 052513 (2002)
- [233] T. Aimi and M. Imada, J. Phys. Soc. Japan **76**, 113708 (2007)
- [234] D. Sénéchal, P.-L. Lavertu, M.-A. Marois, and A.-M.S. Tremblay, Phys. Rev. Lett. **94**, 156404 (2005)
- [235] M. Aichhorn, E. Arrigoni, M. Potthoff, and W. Hanke, Phys. Rev. B **74**, 024508 (2006)
- [236] M. Aichhorn, E. Arrigoni, Z.B. Huang, and W. Hanke, Phys. Rev. Lett. **99**, 257002 (2007)
- [237] K.K. Gomes, A.N. Pasupathy, A. Pushp, C. Parker, S. Ono, Y. Ando, G. Gu, and A. Yazdani, J. of Phys. Chem. Solids **69**, 3034 (2008)
- [238] A. Pushp, C.V. Parker, A.N. Pasupathy, K.K. Gomes, S. Ono, J. Wen, Z. Xu, G. Gu, and A. Yazdani, Science **324**, 1689 (2009)
- [239] A. Dubroka, M. Rössle, K.W. Kim, V.K. Malik, D. Munzar, D.N. Basov, A.A. Schafgans, S.J. Moon, C.T. Lin, D. Haug, V. Hinkov, B. Keimer, T. Wolf, J.G. Storey, J.L. Tallon, and C. Bernhard, Phys. Rev. Lett. **106**, 047006 (2011)
- [240] T. Kondo, Y. Hamaya, A.D. Palczewski, T. Takeuchi, J.S. Wen, Z.J. Xu, G. Gu, J. Schmalian, and A. Kaminski, Nat Phys **7**, 21 (2011)
- [241] P.M.C. Rourke, I. Mouzopoulou, X. Xu, C. Panagopoulos, Y. Wang, B. Vignolle, C. Proust, E.V. Kurganova, U. Zeitler, Y. Tanabe, T. Adachi, Y. Koike, and N.E. Hussey, Nat Phys **7**, 455 (2011)
- [242] E. Morenzoni, B.M. Wojek, A. Suter, T. Prokscha, G. Logvenov, and I. Božović, Nat. Commun. **2**, 272 (2011)
- [243] V.J. Emery and S.A. Kivelson, Phys. Rev. Lett. **74**, 3253 (1995)
- [244] V.J. Emery and S.A. Kivelson, Nature **374**, 434 (1995)
- [245] D. Podolsky, S. Raghu, and A. Vishwanath, Phys. Rev. Lett. **99**, 117004 (2007)
- [246] Z. Tesanovic, Nat. Phys. **4**, 408 (2008)
- [247] I. Ussishkin, S.L. Sondhi, and D.A. Huse, Phys. Rev. Lett. **89**, 287001 (2002)

- [248] E. Fradkin, S.A. Kivelson, M.J. Lawler, J.P. Eisenstein, and A.P. Mackenzie, Annual Review of Condensed Matter Physics **1**, 153 (2010)
- [249] F. Rullier-Albenque, H. Alloul, F. Balakirev, and C. Proust, EPL **81**, 37008 (2008)
- [250] H. Alloul, J. Bobroff, M. Gabay, and P.J. Hirschfeld, Rev. Mod. Phys. **81**, 45 (2009)
- [251] H. Kino and H. Fukuyama: J. Phys. Soc. Japan **65**, 2158 (1996)
- [252] Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, Phys. Rev. Lett. **91**, 107001 (2003)
- [253] B.J. Powell and R.H. McKenzie, Rep. Prog. Phys. **74**, 056501 (2011)
- [254] P. Sahebsara and D. Sénéchal, Phys. Rev. Lett. **97**, 257004 (2006)
- [255] B. Kyung, G. Kotliar, and A.-M.S. Tremblay, Phys. Rev. B **73**, 205106 (2006)
- [256] S. Dayal, R.T. Clay, and S. Mazumdar, Phys. Rev. B **85**, 165141 (2012)
- [257] V.J. Emery, Phys. Rev. Lett. **58**, 2794 (1987)
- [258] A. Macridin, T.A. Maier, M.S. Jarrell, and G.A. Sawatzky, Phys. Rev. B **71**, 134527 (2005)
- [259] W. Hanke, M. Kiesel, M. Aichhorn, S. Brehm, and E. Arrigoni, Eur. Phys. J. **188**, 15 (2010)

11 Superconductivity: 2D Physics, Unknown Mechanisms, Current Puzzles

Warren E. Pickett

Department of Physics, University of California Davis
Davis CA 95616, USA

Contents

1	Overview: Why this topic? What kinds of correlated states?	2
2	Very basic theoretical background	3
2.1	Weak coupling	4
2.2	Strong coupling	6
3	Doped 2D ionic insulators: general aspects	7
3.1	A broad view of the theoretical challenge	7
3.2	Density of states $N(E)$; generalized susceptibility $\chi(q)$	7
3.3	Electronic screening by a sparse electron gas	8
3.4	Dynamics of the coupled ion-electron system	8
4	Electron-phonon coupling in 2D HTS metal MgB₂ class	9
4.1	The surprise of MgB ₂	9
4.2	Superconductor design: attempts within the MgB ₂ class	9
5	Doped 2D ionic insulators: examples	13
5.1	Transition metal nitridochlorides HfNCl and ZrNCl	13
5.2	The TiNCl sister compound	14
5.3	Overview of the transition metal nitridohalides	15
5.4	Related classes of materials	16
6	Transition metal dichalcogenides and oxides: a class, or individuals?	17
6.1	Li _x NbO ₂ : a triangular lattice, single band correlated superconductor	17
6.2	Na _x CoO ₂	20
6.3	Doped transition metal dichalcogenides; recently Cu _x TiSe ₂	20
7	NaAlSi: unusual self-doped semimetallic superconductor	21
8	Doped hydrocarbons: organic crystals	23
9	Summary of main points	24

1 Overview: Why this topic? What kinds of correlated states?

Superconductivity has been beguiling and bedeviling physicists for a century while numerous other quaint and curious collective phenomena have been discovered and analyzed, yet it maintains its mystery in spite of the enormous amount that has been learned and the vast competition for the physical scientist's attention and devotion. Levitation of a magnet over a superconductor that can be turned on and off by anyone using a liter of liquid N₂ to vary the temperature around T_c , fascinates the viewing public (and often us practitioners as well). The "mysteries" of superconductivity – why do the high T_c members have such impressive behavior; can T_c be elevated closer to, or even above room temperature – combine with the unfilled promises of applications of room temperature superconductors to keep this area of study alive in the minds and the laboratories of a large number of scientists.

Though it may surprise the reader, there will be no description or analysis in this lecture of either the high temperature superconducting (HTS) cuprates, with T_c of 130+ K increasing to 160+ K under pressure, nor of the more recent Fe-based HTS superconductors (FeSCs) with T_c as high as 56 K. Neither will heavy fermion superconductivity be discussed. The emphasis in this overview will be on demonstrating that there are other classes of superconductors that are perplexing: their pairing mechanisms are not understood but seem different from the heavily studied classes, hence they are candidates to lead to new classes of high temperature superconductors. The focus here will be a much quieter area of superconducting materials and associated phenomena: strongly two dimensional (2D) band insulators where doping leads to superconductivity and, at least for several, the mechanism seems unrelated to magnetism. We will ask: what are the peculiar superconductors that beg for explanation; what are the characteristics that set them apart from other classes; what types of electron-electron (and electron-ion) correlations determine their behavior? An additional reason for this choice of emphasis is that, after more than 25 years of intense study of the cuprate HTSs and a huge amount of publications, a brief overview would serve little purpose. The intense study of the FeSCs is ongoing, involving numerous issues, and one should have similar reservations about attempting a brief overview.

The pairing mechanism in the cuprate and iron-based HTSs must be magnetism-related, due to the evident competition between magnetic order and superconductivity. Pinpointing the mechanism and why the cuprate and pnictide structures and characteristics are so special (HTS, with T_c above 50 K, occurs only in these two classes), remains one of the outstanding theoretical conundrums in materials physics. The Babel-istic situation was illuminated by Scalapino in his synopsis of the Materials and Mechanism of Superconductivity M²S-HTSC conference (Dresden, 2006). He noted that, by his compilation, the "mechanisms" discussed *at that conference alone* included: (a) Jahn-Teller bipolarons; (b) central role of inhomogeneity; (c) electron-phonon+ U ; (d) spin fluctuations; (e) charge fluctuations; (f) electric quadrupole fluctuations; (g) loop current fluctuations; (h) d density wave; (i) competing phases; (j) Pomeranchuk instabilities; (k) d - d electronic modes; (l) RVB-Gutzwiller projected BCS. Learning more about these terminologies is left to the interested researcher, but it is clear that there is a profusion of concepts and a paucity of consensus on the microscopic mechanism of pairing in cuprate HTS.

New Materials: Recent Superconductor Discoveries

Courtesy:
G. W. Crabtree

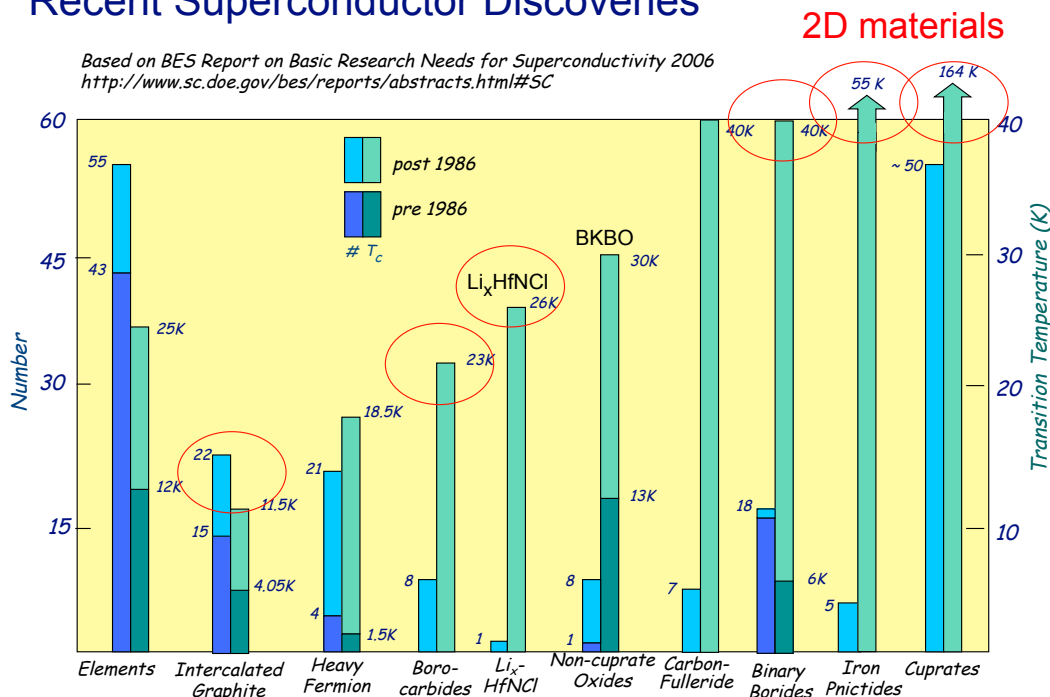


Fig. 1: Plot of data relating to the main classes of high temperature superconductors. Note the legend (upper part of figure): for each class, # means the number of members, T_c gives the maximum critical temperature; bottom part of bar indicates pre-1986 (HTSC breakthrough), upper part of bar indicates post-1986. The main relevance for this paper is emphasized by the red ellipses, which identify the classes with substantial 2D character. The borocarbides have been included as 2D because of the strong layered aspect of their crystal structures; they are however 3D metals. Most of the classes of HTSs are quasi-2D, without any clear connections between most of the classes. Courtesy of George Crabtree; based on data available in 2006.

2 Very basic theoretical background

Practically all theories of superconductivity draw on the basic, Nobel Prize winning theory of Bardeen, Cooper, and Schrieffer (BCS theory [1]). They presumed that there is some effective attraction between electrons (for them, it was due to exchange of virtual phonons, though such details were peripheral) that provided the opportunity for Cooper pairs [2] to form and to spontaneously condense into a collective non-Fermi liquid state – the superconducting condensate – in which these pairs become correlated into a coherent many-body phase. Cooper had just demonstrated [2] that the Fermi liquid ground state is unstable toward the formation of a single such pair, even if the pairing strength is arbitrarily small. A reading of the BCS paper [1] is a must for any student of physics who wishes to acquire a basic understanding of the superconducting state, the spectra, and the low energy, low temperature (T) thermodynamics. The diligent student should even work her way through at least the first ten pages or so of the algebra – it tremendously helps understanding to know something about how the processes are described algebraically.

2.1 Weak coupling

In BCS theory, there is an electronic density of states $N(0)$ at the Fermi level, presumed to vary slowly on the scale of the energy of the virtual boson that transmits the interaction (viz. phonon, in conventional superconductors). The attractive effective interaction $-V$ ($V > 0$) is presumed to be constant up to a cutoff $\hbar\omega_c$ of the order of a phonon energy. The approach used by BCS was to guess the form of a correlated ground state wavefunction depending simply on few parameters, and obtain the parameters via a mean-field minimization of the energy, first at $T = 0$ and then at finite temperature. Conventional SCs, by the way, are exceedingly good examples of “mean field transition” systems, the critical region around T_c being exponentially small and unobservable. In the weak coupling limit, and only in that limit, T_c is exponentially related to coupling strength λ :

$$k_B T_c = 1.14 \hbar\omega_c e^{-1/\lambda} \quad (1)$$

with

$$\lambda = N(0)V. \quad (2)$$

At moderate $\lambda \sim 0.5$ - 0.75 , say, the equation must be solved numerically for T_c , and $T_c(\lambda)$ is quasilinear rather than exponential. The strong coupling regime must be treated separately and is discussed below.

The superconducting gap Δ_k is constant over the nondescript Fermi surface (FS) in the broadest form of BCS theory. The more general expression for the momentum dependence of Δ_k over a general FS is given by [3]

$$\Delta_k = - \sum_{k'} V_{k,k'} \frac{\Delta_{k'}}{2\sqrt{\varepsilon_{k'}^2 + \Delta_{k'}^2}} \tanh \frac{\sqrt{\varepsilon_{k'}^2 + \Delta_{k'}^2}}{2k_B T} \quad (3)$$

$$\rightarrow - \sum_{k'}^{FS} V_{k,k'} \frac{\Delta_{k'}}{2\sqrt{\varepsilon_{k'}^2 + \Delta_{k'}^2}}, \quad (4)$$

where the last expression is the $T \rightarrow 0$ expression. More generally, band indices are also required. The $\tanh()$ term arises from Fermi-Dirac thermal distribution factors, ε_k is the non-interacting band energy, and $V_{k',k}$ is the matrix element for scattering of pairs between k and k' on the Fermi surface. The critical temperature T_c is determined from the linearized gap equation in the limit of $\Delta_k \rightarrow 0$:

$$\Delta_k = - \sum_{k'}^{FS} V_{k,k'} \frac{\Delta_{k'}}{2\varepsilon_{k'}} \tanh \frac{\varepsilon_{k'}}{2k_B T} \quad (5)$$

T_c is the highest temperature for which there is a nonvanishing solution for Δ_k .

Although these gap equations (finite T and the linearized version giving T_c) are in the weak coupling limit (and subject to other simplifications made in BCS theory), they are very commonly applied, or at least cited, in situations where they have not been justified. This reflects the confidence that theorists have that some “essence” of pairing superconductivity is contained in these equations. The linearized equation (5) is especially prevalent in modern discussions. With the discovery of the HTS cuprates, there quickly arose a great deal of interest in Fermi

surface nesting, in hot spots on the Fermi surface (van Hove singularities extending toward infinity), and in pairing interactions in which there is *strong anisotropy*. This anisotropy usually does not become important for electron-phonon pairing, but if one assumes that magnons can induce pairing analogously to phonons – the virtual boson that is exchanged is a magnon rather than a phonon – then anisotropy becomes paramount. In the cuprates, it is presumed that (i) the strong Coulomb repulsion U on Cu keeps potentially pairing electrons off the same Cu site, and (ii) the interaction is dominated by strong short-range AFM fluctuations.

The influence of the linearized gap equation Eq. (3) on contemporary superconductivity theory can hardly be overemphasized. In BCS theory the (maximally isotropic) coupling must be *attractive* to obtain solutions of the gap equation. When the interaction is anisotropic and even repulsive on average, gap solutions (*i.e.* superconducting states) for non-zero Δ_k can still be obtained. The change of sign with angle of the pairing interaction can be compensated by a change in sign of Δ_k with angle. The quantity under the integral in Eqs. (3) and (5) is then predominantly of one sign, as is the case for an attractive isotropic coupling and isotropic gap. The interested student or postdoc will benefit in understanding when studying the above gap equations by expanding the various functions in spherical (3D) or circular (2D) harmonics, and making reasonable assumptions about the behavior with the magnitude of $|k - k_F|$ (constant up to a cutoff, say). Looking at the linearized case $T \rightarrow T_c$ and contributions from the Fermi surface (FS) $\varepsilon_k \rightarrow 0$ are most useful. Expressions for general FSs (number and shape) can be written down using the Fermi surface harmonics of Allen [4], although all but the simplest situations will require numerical solution.

These anisotropic gap solutions are a realization of the “theorem” of Kohn and Luttinger [5], which pointed out that such anisotropic solutions would exist for anisotropic coupling, although at the time they were expected to have implausibly small values of T_c . Such *exotic pairing*, or *exotic order parameter*, has become over the past two decades commonplace in theories and have found strong confirmation in cuprates and some heavy fermion SCs. Several experiments demonstrate (or strongly imply) that the hole-doped cuprates have a *d*-wave order parameter (gap function Δ_k), with angular dependence like $\text{sgn}(x^2 - y^2)$; thus it is referred to as a $d_{x^2-y^2}$ (angular momentum of the pair $\ell = 2$) symmetry order parameter for the superconducting state. In heavy fermion superconductors, $\ell = 3$ pairing seems likely [6] in UPt₃, which has hexagonal symmetry that conspires against *d*-wave symmetry. In considering new superconductors, one of the most valued characteristics is determining if pairing is “conventional” (isotropic $\ell = 0$) or “exotic” (anisotropic $\ell > 0$), because this character is likely to reflect conventional electron-phonon or unconventional pairing, respectively. In the latter case the gap (usually) has nodes on the Fermi surface, hence there is no true gap in the superconductor’s excitation spectrum. This aspect impacts thermodynamics strongly, and the temperature dependence of thermodynamic quantities as $T \rightarrow 0$ is the most common evidence quoted for the (an)isotropy of pairing. If there is a gap, the heat capacity goes exponentially to zero as $T \rightarrow 0$, if not it approaches zero as a power law in temperature.

2.2 Strong coupling

Strong coupling indicates the regime $\lambda \geq 1$ where perturbation theory in the electron-phonon coupling strength no longer holds, and many aspects of the physics are different. For the phonon mechanism, the generalization of Eliashberg [7] of BCS theory to strongly coupled electron-phonon models was extended into an extremely detailed and strongly nuanced, material-specific formalism by Scalapino, Schrieffer, and Wilkins [8]. It is this generalization that is now commonly referred to as strong coupling *Eliashberg theory* (as opposed to weak coupling BCS theory). Together with the introduction, at the same time, of density functional theory (DFT) by Hohenberg, Kohn, and Sham [9, 10] and its subsequent very extensive development, DFT-based Eliashberg theory has been shown repeatedly to describe phonon-paired superconductors quite reliably. The primary restriction for the applicability of Eliashberg theory is that the effective Coulomb repulsion between electrons is retarded in time and is weak, and thus can be characterized by a repulsive effective interaction strength $\mu^* = 0.1 - 0.2$. When the impact of the Coulomb interaction is great, which usually manifests itself in magnetic behavior, no justifiable theory of superconductivity exists.

Within DFT-Eliashberg theory the electron-phonon interaction (EPI) strength is given for an elemental metal by

$$\lambda = \frac{N(0)I^2}{M\langle\omega^2\rangle} \equiv \frac{K_e}{K_\ell}; \quad I^2 = \langle\langle V_{k,k'}^2 \rangle\rangle. \quad (6)$$

Here $V_{k,k'}$ is matrix element for scattering from the FS (k) to the FS (k') by an atomic displacement, M is the ionic mass, and the phonon frequency average is weighted appropriately by matrix elements. This expression is precisely true for elemental SCs (note that only one mass enters) but survives as a guideline for compounds where the character of coupling can be much richer. This form emphasizes that λ represents the ratio of an “electronic stiffness” K_e and the (textbook) lattice stiffness K_ℓ , i.e. determined by the interatomic force constants. For a harmonic lattice the product $M\langle\omega^2\rangle$ is independent of mass, so the mass dependence of T_c comes solely from the prefactor $\omega_c \propto 1/\sqrt{M}$. This mass dependence reflects a crucial factor in EPI-based pairing that has been recognized and exploited since the prediction that metallic hydrogen should be a room temperature superconductor: other factors i.e. the electronic structure, being the same, materials with lighter ions should have higher T_c simply because the fundamental energy scale ω_c is higher.

Strong coupling Eliashberg theory is much richer than BCS theory. Allen and Dynes [11] analyzed materials trends and the Eliashberg integral equation for T_c and demonstrated, among other results, that at large coupling $T_c \propto \sqrt{\lambda}$ and thus is unbounded, providing strong encouragement for the likelihood of (much) higher T_c SCs. Since this lecture will not deal with issues of strong coupling, this and other aspects of Eliashberg theory are not needed for the discussion and will not be presented here.

3 Doped 2D ionic insulators: general aspects

3.1 A broad view of the theoretical challenge

Conventional pairing, that is, electron-phonon, is fully attractive: every phonon contributes to an *attractive* interaction between electrons on the Fermi surface. This attraction also operates in unconventionally paired superconductors (such as the HTS cuprates) but seems to be ineffective. It may even be detrimental to the eventual superconducting state if the gap symmetry is exotic. The always attractive electron-phonon interaction strongly favors a fully symmetric gap, although highly anisotropic pairing and complex FSs might provide more interesting order parameter symmetry. The screened Coulomb interaction between electrons is *almost always repulsive*, but the average repulsion may become irrelevant if pairing is anisotropic, according to current understanding, and the Kohn-Luttinger result mentioned earlier.

The purpose of this lecture is to provide an overview of a few classes of materials – 2D doped ionic insulators – wherein the Coulomb interaction between electrons, normally repulsive, may acquire new traits beyond what have been studied thoroughly so far. Two dimensionality may play a special role [12], through the phase space that it carries and the manner in which interactions are shaped. A relatively low density of doped-in carriers may introduce novel dynamical effects. And when these two aspects are present in the background of highly charged, vibrating ions, the underlying behavior may include unusual emergent aspects (to use a term much in fashion these days).

3.2 Density of states $N(E)$; generalized susceptibility $\chi(q)$

Near a band edge with normal quadratic dispersion in 2D, the density of states $N(E)$ is a step function. Thus small doping leads already to a large, metallic value of $N(0)$ unlike in 3D semimetals where $N(0)$ increases monotonically with the carrier concentration and may be arbitrarily small at low doping. It is for this reason that B-doped diamond becomes superconducting but with only a modest value of T_c (up to 11 K has been reported [13]). In other respects this system is in the MgB_2 class (more specifically, the hole-doped LiBC) class, where electron-phonon matrix elements are large and the relevant phonon frequency is very large.

Due to the 2D character, the FSs are closed curves versus the closed surfaces that occur in 3D. Near a band edge these will be circles or nearly so, making their algebraic description and even that of the complex generalized susceptibility $\chi(Q, \omega)$ possible [14]. Thus the underlying mean-field, static lattice electronic structure and linear response is straightforward, even simple, to model. $\chi^{2D}(Q, \omega)$ is available analytically for a single circular FS, and for a few symmetry-related FSs, which comprise a multi-valley system, this χ^{2D} for a single band will be supplemented by a sum of inter-FS terms $\chi^{2D}(|Q - Q_s|, \omega)$, the same form but for initial and final FSs separated by the spanning wavevector(s) Q_s . (If band extrema do not occur at high symmetry points, the FSs may be ellipsoidal instead of circular and the form of $\chi(Q, \omega)$ becomes anisotropic and correspondingly more involved.

3.3 Electronic screening by a sparse electron gas

Due to the electronics applications of 2D electron gases (2DEGs), their dynamical response has been studied extensively. Within the random phase approximation (RPA), the plasmon dispersion is given implicitly by $\epsilon(Q, \omega_p) = 1 - v(Q)\chi(Q, \omega_p) = 0$ where $v(Q)$ is the unscreened Coulomb repulsion; that is, a “response” can occur in the absence of any perturbing potential when the screening ϵ^{-1} diverges. Whereas the usual long wavelength plasmon in 3D behaves as $\omega_p^2(Q) = \omega_p^2(0) + BQ^2 + \dots$, in 2D the much stronger dispersion $\omega_p = A\sqrt{Q} + \dots$ holds. The plasmon vanishes at $Q = 0$, leading to strong dynamical behavior (screening, perhaps over-screening, or other unconventional behavior) in at least a small region around $Q = 0$. The idea that 2DEGs might be fertile ground for superconductivity has been around for some time [12]. Ionic insulators have high frequency transverse (TO) and longitudinal (LO) optical modes also around $Q = 0$ that will cross and interact with the plasmon, leading to coupled modes that are candidates for unconventional dynamical behavior and possible pairing of electrons. Layered crystals do not present strict 2DEGs however; they are instead (naturally occurring, or nowadays sometimes grown atomic layer by layer) multilayers. In the multilayer case the Coulomb interaction couples the response of neighboring layers (even in the absence of electron hopping between layers) and the $Q \rightarrow 0$ plasmon remains finite [15] but may still be very soft and strongly coupled to ionic dynamics.

3.4 Dynamics of the coupled ion-electron system

Allen, Cohen, and Penn [16] (ACP) have emphasized that the total interaction between two electrons in a crystal involves the combined dynamic polarizability (i.e. the total dielectric function) of the electronic system and the lattice, and they have provided a firm background for the study of such systems. When the conduction electron density is low, the competition between weak dynamically screened repulsive (electron-electron, cation-cation, anion-anion, electron-anion) and attractive (electron-cation, anion-cation) interactions may produce new “regions” of effective attraction. They derived within a general formalism that can be approached in a material-specific, first principles way (such as by using a DFT starting point) that, even taking into account interactions between electrons, between ions, and between ions and electrons, the polarizability of the system is the sum of two terms: that of the vibrating ions, and that resulting between electrons interacting through the dynamically screened Coulomb interaction.

Bill and collaborators [17, 18] have constructed the most detailed model of 2D superconductivity arising from coupled phonon-plasmon modes, giving particular attention to special aspects of plasmonics in two dimensions. Related themes appear occasionally elsewhere in the literature, for example that of Askerzade and Tanatar [19] and of Falter and coworkers [20, 21]. Other unconventional interaction channels may arise in such systems. Ashcroft has emphasized polarization waves due to flexible semicore electrons [22] as possibly contributing to pairing. In 2D lattices where there is a natural axis (the c axis), polarization modes (“ferroelectric fluctuations”) may have more impact than in 3D lattices.

4 Electron-phonon coupling in 2D HTS metal MgB_2 class

One focus of this lecture is two dimensionality and its relation to superconductivity and T_c , so it is important to review (albeit briefly) the spectacular surprise presented by the discovery of superconductivity in MgB_2 by Akimitsu's group in 2001 [23]. The account of the quest for other MgB_2 -like materials, following in Sec. 4.2, is both intriguing and sobering.

4.1 The surprise of MgB_2

MgB_2 , as a standard *sp* HTS metal, has $T_c = 40$ K when the light isotope of B is used. It is described well by Eliashberg theory (in its multiband extension) as implemented in DFT-based electron-phonon calculations [24–28]. It provided many lessons by violating nearly all of the conventional wisdom of the time: (a) it is an *sp*, not *d* metal; (b) it is strongly 2D rather than 3D; (c) it becomes a HTS superconductor due to *extremely* strong coupling to *extremely* few (3%) of the phonons, rather than having the strength spread rather uniformly over the phonon spectrum. It is best regarded not a standard metal, but as a self-doped semimetal; the crucial σ -bonding band is nearly filled. The basic aspects of the electronic structure and coupling – that high frequency B-B stretch modes are extremely strongly coupled to the strongly bonding B-B states at the Fermi surface – can be well understood in terms of simple formal expressions, which provides an explicit recipe [14, 29] for the type of extension from MgB_2 that could provide much higher T_c within this class of metal. The concept is provided briefly in Fig. 2 and its caption. Simply put, change the Fermi surfaces to make use of coupling to more phonon modes and provide a larger electronic density of states, while retaining the structure that gives very strong bonding (large electron-phonon matrix elements).

4.2 Superconductor design: attempts within the MgB_2 class

The simplicity of the crucial features of MgB_2 has encouraged discovery or design of additional members of this class of superconductor, as described briefly in this subsection.

4.2.1 Hole-doped LiBC

The first such proposed extension fits in with the focus of this lecture in many respects, except for the fact that electron-phonon coupling is not really a focus. LiBC is isostructural and “isovalent” with MgB_2 (Li having one less electron than Mg, C having one more electron than B), but it is insulating due to the inequivalence of B and C on the honeycomb sublattice. Hole-doping in this covalent/ionic insulator by partial removal of Li, while retaining the crystal structure and obtaining a black (likely conducting) sample, was reported by Wörle *et al.* [30] in 1995. Calculations of the electron-phonon coupling strength by Rosner *et al.* predicted that such doping would lead to T_c of 75 K or higher [31, 32]. Li_{1-x}BC is a MgB_2 look-alike system, with the increase in T_c over that of MgB_2 resulting from the stronger B-C bonding compared to B-B bonding in MgB_2 , giving both larger matrix elements and a higher phonon energy scale. Work

Design of higher T_c superconductors: is it viable?

Rational Design/Search for new hTS

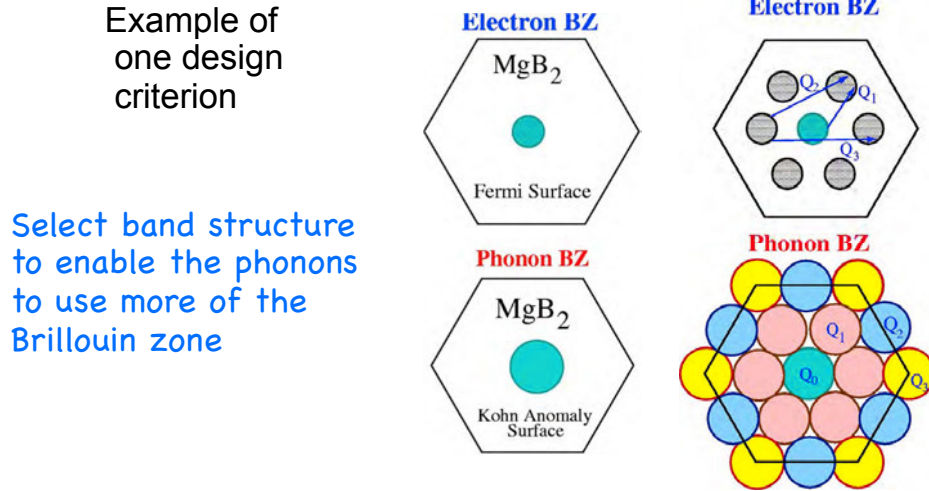


Fig. 2: A proposal for rational design of higher T_c materials. Rational design is possible because the electron-phonon interaction and resulting superconductivity is extremely well understood (in weakly correlated Fermi liquid metals). MgB_2 makes use of extremely strong coupling $\lambda_{qv}=20-25$ to only 2-3% of the phonon modes: the two bond stretch modes polarized in the plane (hence 2 of 9 branches) in only 8% of the zone (where $q < 2k_F$). Adding Fermi surfaces in other parts of the zone provides coupling from branches at other values of q : Q_1, Q_2, Q_3 connecting the various sheets of Fermi surface. A nearly ideal scenario is pictured on the two zone figures on the right. See Refs [14, 29] for further description.

on this system, with the most extensive being done by the Rosseinsky group [33], indicates unfortunately that the doped system is prone to (structural and phase separation) instabilities that prevent realization of the desired phase. The report of Wörle *et al.* has never been confirmed. Lazicki and collaborators [34] pursued the possibility that pressure might close the gap and induce metallization and thereby superconductivity. The structure remained stable to 60 GPa, and density functional calculations predicted that metallization in this structure would not occur until at least 345 GPa. This system illustrates how higher T_c is in practice often limited by instabilities that can appear in assorted flavors, while the underlying theory provides no upper limit [11] on the possible value of T_c .

4.2.2 Transforming graphite into pseudo-MgB₂

Simultaneously with the study of Li_{1-x}BC , our group considered a different means to obtain a MgB_2 -like material. MgB_2 is, after all, graphite with an extra three dimensional band in the background. The difference is that MgB_2 has a different potential between the honeycomb layer and the interstitial, or Mg, layer. The σ -bonding band is present in graphite, but its upper edge is 2 eV below the Fermi level, which is determined by the positioning of the π -bonding band at

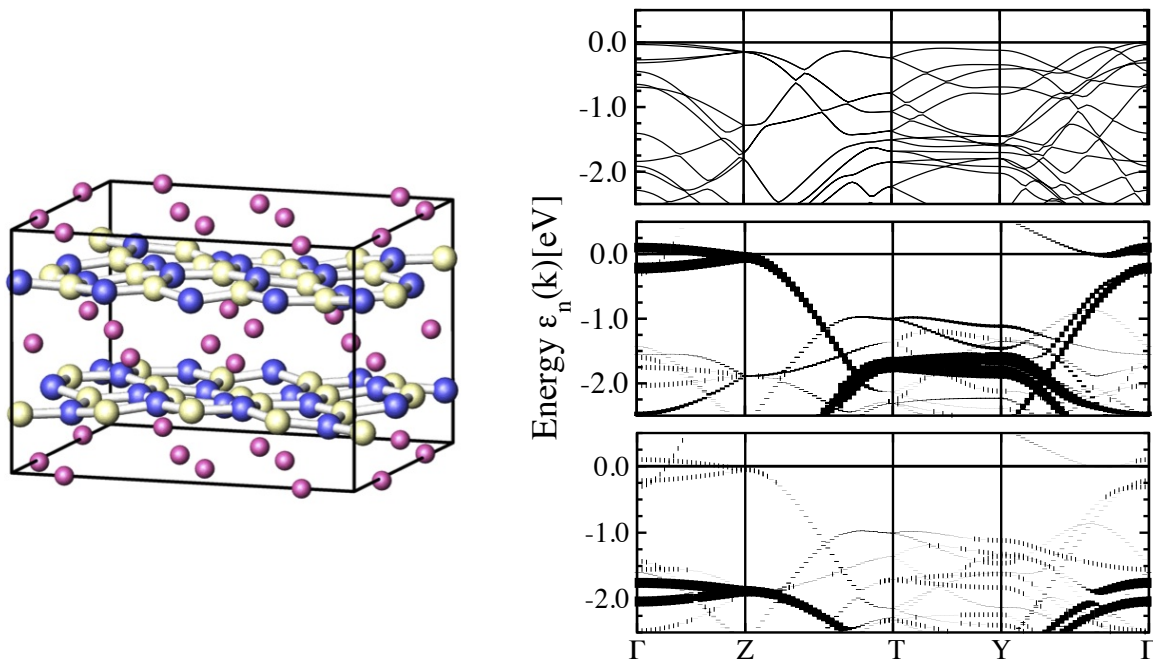


Fig. 3: Left panel: the orthorhombic crystal structure of MgB_2C_2 in perspective view. Mg, B and C are represented by pink (small spheres) blue, and light yellow (larger) spheres, respectively. The buckling of the visualized covalent B-C bond causing the deviation from the ideal hexagonal plane is visible. Right panel: Band structure for the undistorted (upper subpanel) and B-C σ 'fat bands' for a frozen-in bond stretching phonon (middle and lower subpanels) with a bond elongation around the rms value. The large energy difference between the $2p_x$ orbitals (middle panel) and the $2p_y$ orbitals (lower panel) indicates the very strong deformation potential for this mode. From H. Rosner, A. Kitaigorodsky, and W.E. Pickett, unpublished.

symmetry point K in the Brillouin zone, which has in the meantime asserted its infamy in the plethora of graphene research of the past decade. Our idea was simple – in hindsight, it was simplistic. What seemed to be necessary was to lower the Fermi level by 2 eV in graphene. This could be done by intercalating it with a highly electronegative ion. The most electronegative one, and also a small one, is fluorine. Joonhee An [35] carried out the calculation of FC_2 in the MgB_2 structure. Fluorine did become a negative ion of course, but another change that we had not anticipated was a shift in Madelung potential. This shift counteracted to a great degree the charge transfer, and left the Fermi level well away from the σ -bonding bands. Thus this approach to HTS design did not work.

4.2.3 Hole-doped MgB_2C_2

The (unsuccessful) example of LiBC has encouraged further exploration into this direction of finding MgB_2 -like materials. The borocarbide compound MgB_2C_2 is isovalent, and structurally similar, to the (super)conductor MgB_2 and to insulating LiBC. The structure [36] is pictured in Fig. 3. Due to the placement of Mg ions, the honeycomb B-C layers are dimpled somewhat. Density functional based electronic structure calculations and electron-phonon coupling

strength calculations [37–41] show that MgB_2C_2 (i) is insulating like LiBC due to the modulation of the honeycomb B-C layers, (ii) exhibits a rather high B-C σ -band density of states close to the Fermi level when slightly hole doped, and (iii) shows a strong deformation potential with respect to the B-C bond stretching modes, as demonstrated in Fig. 3. If large enough hole doping of the system can be achieved, such as by replacement of Mg by Li, it should be superconducting at temperatures comparable to MgB_2 . Mori and Takayama-Muromachi reported attempts to hole dope this compound both on the Mg site and within the B-C network, but found no indication that the dopants actually entered the crystal structure [42]. Yan and collaborators [43] have recently provided an extensive density functional study of the structural and thermodynamic properties of insulating MgB_2C_2 .

4.2.4 Hole-doped BeB_2C_2

Forty years after the compound was first synthesized, the structure of BeB_2C_2 was finally solved by Hoffman *et al.* [44]. While possessing the same honeycomb B-C layers as LiBC and MgB_2C_2 , it has a specific, non-intuitive stacking due to the position of the interlayer Be, which likes to coordinate on one side with a single C atom. Before the structure was known, Moudden calculated the electronic structure and electron-phonon coupling strength for a different structure also based on the B-C honeycomb rings [45]. Although the bands at the bottom of the gap, which are the active states when hole-doped, are considerably more intricate than MgB_2C_2 , Moudden obtained a larger coupling strength and T_c than for MgB_2 , for the same reasons as for hole-doped LiBC and MgB_2C_2 . In our unpublished work [46] using the experimental structure, we find that the relevant bands are simpler than those obtained by Moudden, and in fact much more MgB_2 -like. Not surprisingly, we also obtain very strong coupling to the B-C stretch modes and a probable T_c higher than in MgB_2 .

4.2.5 Comments on this class of doped insulators

MgB_2 has spawned the study of these hole-doped ABC and AeB_2C_2 insulators (A = alkali; Ae = alkaline earth), which has led to theoretical predictions of high temperature superconductivity. This activity has been disappointingly unproductive so far: in the cases attempted experimentally progress has been stymied by the inability to introduce the dopant (or the vacancy) in a random alloy fashion as presumed by the theory – chemistry gets in the way. A further member, CaB_2C_2 , also exists [47]. It also sports a 2D B-C network. Its structure is, however, composed of B-C octagons and B-C diamonds rather than the honeycomb network of the others. This system certainly seems worthy of study and attempts at doping.

This progression from the 2D B-B net of MgB_2 to the B-C nets of the Ae compounds can be taken a step further: the network components can be changed from B-C to Be-N, substantially increasing their distance in the periodic table and thereby reducing the degree of covalency while retaining the overall isovalent nature. The compounds are indeed insulating, and $Ae = \text{Ca}, \text{Sr}, \text{Ba}$ have the same structure as CaB_2C_2 . The $Ae = \text{Mg}$ compound has a distinctive structure built on a strongly puckered Be-N bilayer with Mg ions distributed between the layers [48].

5 Doped 2D ionic insulators: examples

5.1 Transition metal nitridochlorides HfNCl and ZrNCl

The class $\mathcal{T}\text{NCl}$, where \mathcal{T} is a group IV-B transition metal Hf, Zr, and Ti (there are also a few Br (instead of Cl) members) is the prime example of the type of superconductor that will receive emphasis in the remainder of this lecture. These three isovalent compounds display T_c up to 25.5, 15.5, and 17.5 K, respectively, when electron doped [49–54]. None have been hole doped, and the active states in that case would be very different, being N $2p$ states. Undoped, they are highly ionic $\mathcal{T}^{4+}\text{N}^{3-}\text{Cl}^+$, moderate gap ($\sim 2\text{--}3$ eV) insulators with strongly layered structures. The $\mathcal{T} = \text{Hf}$ and Zr members, which are extremely similar in electronic structure, have a somewhat dimpled BN-like alternating honeycomb -Hf-N-Hf-N- bilayer with Hf-N bonds coupling the bilayers. Cl caps the honeycomb holes above and below, resulting in neutral layers that are van der Waals bonded.

The covalence between the d and the N $2p$ states in these ionic compounds is evident in their Born effective charges (BECs). The BECs are highly anisotropic, often differing by a factor of two between in-plane and out-of-plane, and in some cases the values are well above (in magnitude) their formal charges [55]. The trend in magnitudes increases from $5d$ to $3d$. It is these values, and the resulting internal electric fields, that a low density of carriers will experience.

Alkali ions, without or with larger organic molecules, can be intercalated between the layers to induce conductivity and superconductivity, and recently it has been shown that (trivalent) rare earths can also be used for the doping, with T_c remaining the same. Once somewhat beyond the insulator-superconductor transition [56], T_c is almost independent of the carrier density (the doping level). The layering and bonding of the TiNCl compound is very similar, although the crystal symmetry is orthorhombic rather than hexagonal/rhombohedral as are the others. We return to TiNCl in the next subsection. Schurz *et al.* [57] have provided a recent experimental overview of these materials, primarily on synthesis and structure.

It was shown by Weht and coauthors [58] that the doped electrons are accommodated at the bottom of the \mathcal{T} d -band, which has substantial in-plane dispersion (the effective mass is of the order of unity) as shown in Fig. 4. Thus due both to the broad d band and the small carrier concentration (far from half filling) the type of strong correlation effects that are ubiquitous in transition metal oxides appear not to be dominant in this transition metal nitride, and there is no experimental evidence of correlation effects such as magnetic moments and magnetic ordering, or orbital or charge ordering, etc. It was also found that the electronic structure of isostructural HfNCl and ZrNCl are extremely similar, yet their transition temperatures have consistently been observed to differ by a factor of two: 25 K versus 12–13 K for ZrNCl over most of the doping range. The difference in T_c is opposite to what would be expected from a BCS isotope effect. It must then be related to other features: to differences in force constants or electronic response (*viz.* Born effective charges, or higher frequency response) though the similar band structures suggest they should not differ much, or to the factor of two difference in \mathcal{T} -mass (178 amu versus 91 amu) that affects lattice polarization.

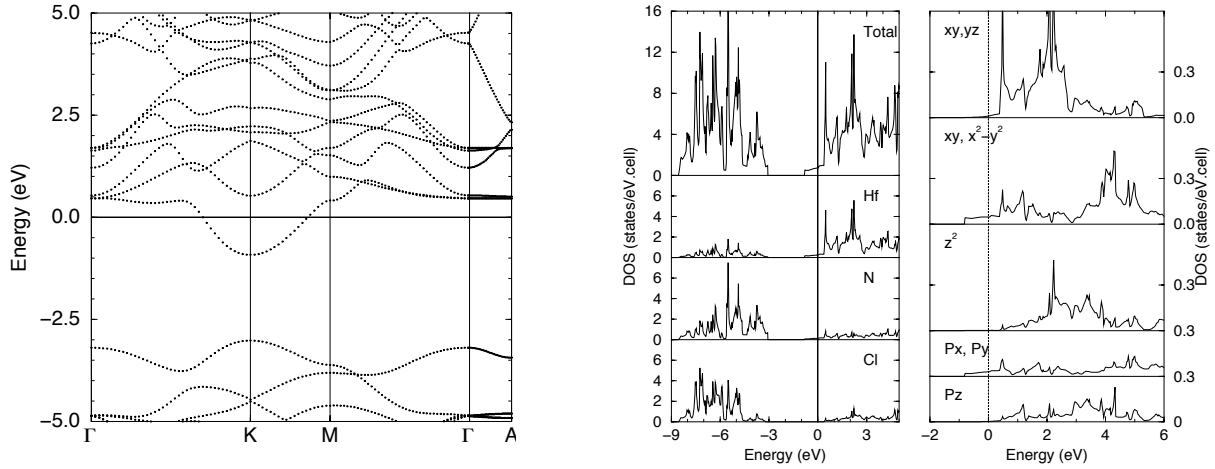


Fig. 4: Left panel: Band structure of $\text{Na}_{0.25}\text{HfNCl}$ along the hexagonal symmetry lines, calculated in the virtual crystal approximation. The five bands above the gap are heavily Hf 5d in character, while the bands below are filled N^{3-} 2p bands. The Cl 3p bands lie somewhat deeper. The flatness along Γ -A indicates the very strong 2D character of the bands of interest. Right panels: The total, atom projected, and 5d and 2p projected densities of states. Note that the Fermi level lies in a region of rather low DOS of d_{xy} , $d_{x^2-y^2}$ character. The other three 5d bands lie ~ 1 eV higher due to ligand field splitting. From Weht *et al.* [58].

Heid and Bohnen [59] carried out density functional linear response calculations of the el-ph coupling strength and character in A_xZrNCl in the same manner as is commonly done for Fermi liquid metals, and found the electron-phonon coupling $\lambda \approx 0.5$ is much too small to account for the observed value of T_c . Akashi *et al.* [60] have provided an application of “DFT for superconductors” to these nitridochlorides, and also conclude that something besides the usual Eliashberg theory is required to understand their superconductivity. This theoretical approach presupposes that (i) carriers are present in a weakly correlated Fermi liquid, and (ii) doping can be treated in the virtual crystal approximation (VCA). The VCA corresponds to adding charge without much concern how it got there and treating it self-consistently, which can be important [61]. It is clear, however, that the materials are more complex than that. The most obvious evidence is from the observation that, in Li_xZrNCl , metallic conduction does not occur until a critical concentration $x_{cr} = 0.06$ is reached [56]. It is noteworthy that $x_{cr} = 0.15$ is different for the doped HfNCl compound [62], however it should be kept in mind that the doping was done in a different manner. The VCA band structure remains, by supposition, that of a conventional Fermi liquid, however, to arbitrarily small doping levels. The interactions that keep the carriers localized at low doping are surely essential to address this class of materials theoretically. For this, the Born effective charges [55] and perhaps nonlinear effects should be important. In addition, $T_c(x)$ in the Zr compound is maximum at x_{crit} [56], about 25% higher than the value of ~ 12 K over most of the measured range of x .

5.2 The TiNCl sister compound

α -TiNCl is an orthorhombic ($Pmmn$, space group 59) member of this class with the FeOCl prototype structure, with similarities and differences when comparing with the nitride halides

in the previous subsection. It is a strongly layered compound including a double layer of Ti-N nets analogous to those of HfNCl and ZrNCl. Each net however has the topology of a single NaCl (001)-layer rather than a honeycomb type, but the layers are displaced so Ti is coordinated with four N ions (two in the same layer, two in the other) and with two Cl layers above (or below). The layers are strongly buckled. The Ti-N bilayer is decoupled electronically from the bilayers above and below, giving it 2D character, but the coordination and bonding are quite distinct from that of ZrNCl and HfNCl. More description, and references, are provided by Yin *et al.* [63].

The band gap of TiNCl is much smaller (0.5 eV) than that of its cousins, and also differs in having the gap occur at Γ rather than at the zone corner. Thus when electron-doped, there is a *single cylindrical Fermi surface* surrounding Γ -Z rather than two at the K and K' points. The character is again in-plane ($3d_{xy}$) and the band is ~ 2 eV wide. The DOS near E_F , $N(0)$, is similar to those of ZrNCl and HfNCl. And, of course, there is the fact that T_c lies in the same, impressively high, range (17 K).

Yin *et al.* [63] assembled a tight binding model Hamiltonian based on Wannier functions and a Hubbard on-site repulsive interaction, and proceeded to calculate the charge and spin susceptibilities $\chi_{ijkl}^{c,s}(\vec{q})$, where the subscripts label the Wannier functions. They presented a few elements that are expected to have the most weight in the full susceptibility, and noted that the approximate square symmetry of the Ti-N bilayer (at least when viewed from above) is strongly broken by some elements of this susceptibility matrix. The calculated Born effective charges are also strongly anisotropic in-plane as well as out-of-plane. The q -dependence of $\chi_{ijkl}^{c,s}(\vec{q})$ will be useful when measurements of the charge and spin fluctuation spectrum are available. They will also be useful if electronic, rather than phononic, pairing mechanisms are considered.

5.3 Overview of the transition metal nitridohalides

It will be instructive to make a list of salient aspects of this class of superconductors, which is part of the broader class of *transition metal pnictide halides* that has been labeled “under explored” [64]. Such a list should contain several clues about the origin of their remarkable superconductivity and more generally about the importance of two dimensionality and doping into ionic insulators.

- The occurrence of superconductivity and the value of T_c is weakly dependent on the type and amount of doping, indicating a robust feature that is insensitive to details such as stacking of successive $(\text{TNCl})_2$ layers, or manner of doping. ZrNCl can even be doped with Cl^- vacancies to superconduct at 12-14 K [65], which is the same range of T_c that arises from alkali atom intercalation.
- In-plane symmetry seems to be of little consequence. The Hf and Zr members have hexagonal, isotropic symmetry in the a - b plane, while the Ti member has strongly anisotropic Born effective charges and susceptibilities with rectangular, i.e., anisotropic symmetry.

- A relatively large critical concentration of carriers ($x_{cr} = 0.06$ for Li_xZrNCl) is required for the insulator-to metal/superconductor transition. At lower concentrations, the doped-in electrons are “solvated” into the transition metal bands, presumably as immobile polarons. This concentration corresponds to two carriers per 4×4 supercell of ZrN bilayers, a low but not truly sparse carrier density. The 2D density parameter is $r_s \sim 20/\sqrt{\epsilon}$, where ϵ is the background dielectric constant of the insulator.
- T_c is maximum in $\text{Li}_{1-x}\text{NCl}$ at the metal-insulator transition, $x_{cr} = 0.06$, as discussed above. This is surely an important clue, given that T_c is so insensitive to other factors (interlayer spacing, type of dopant, doping level). This fact also prompts the question: can x_{cr} be decreased in some (otherwise innocuous) manner, and if so, will T_c continue to rise as the doping level decreases?
- The N isotope effect has been reported [66] to be 0.07 ± 0.04 , reflecting little dependence on N mass. This is quite small and uncertain, but possibly nonzero. Note that the actual shift ΔT_c was 0.06 ± 0.03 K, which is nearing the limit of clear detectability. In any case, the N isotope effect is at least extremely small.
- The factor-of-two difference in T_c between the Hf and Zr compounds invites study; so far there are not even any reasonable speculations on the origin of this difference. The electronic structures are nearly indistinguishable. Interpreted as an isotope effect, (i) the value is very large but also of the wrong sign, and (ii) the difference in T_c is as large as one of the T_c s, so an isotope exponent that assumes $\Delta T_c/T_c$ is small is an inappropriate representation. Since the electron-phonon λ (evaluated in the usual manner) is seemingly small, there is little reason to expect this to be a standard isotope shift anyway.
- TiNCl has an analogous band structure and a metal-nitride bilayer also, and a T_c midway between its two cousins. Yet the lattice symmetry and the Fermi surfaces are different. Supposing the pairing mechanism is the same, these similarities and differences provide clues and potential insight for the microscopic behavior impacting superconductivity.
- The Born effective charges [55] provide the electrodynamic effects of vibrating charged ions in the weakly screened limit. If some correlation can be found between them and superconducting characteristics, it could provide important clues to the pairing mechanism.

5.4 Related classes of materials

One can guess that there must be a substantial number of 2D ionic band insulators that make reasonable candidates for superconductors when doped. We provide some brief comments here. BaHfN_2 seems to be a minor variant of the $\mathcal{T}\text{NCl}$ class. Its structure is analogous, having transition metal nitride (Hf-N) layers bounded by the more ionic layers, which in this case contain BaN [55]. The electronic structure is analogous: N $2p$ states are filled, and the conduction band states that are available for electronic carriers above a (calculated) gap of 0.7 eV are Hf $5d$ states, hybridized with N $2p$ states. This compound differs from the $\mathcal{T}\text{NCl}$ class in a way that

may be important for synthesis: it has only a single reactive anion (N). Many 2D materials are grown layer by layer by sputtering, plasma laser deposition, or molecular beam epitaxy. Much experience has been gained in dealing with multiple cations in the chamber, but usually a single anion is used (and that is almost always oxygen).

The sister compounds SrZrN_2 and SrHfN_2 have also been synthesized [67]. The growth in nitride synthesis has led to an expanding number of transition metal nitrides, many of which have strongly 2D structures of the type of interest here, while other have “low dimensional” or somewhat open structures that are not strictly 2D [64, 68]. Given the impressive superconductivity in the TNCl class, doping these materials may well provide unusual insulator-metal transitions and perhaps some even more impressive superconductors.

6 Transition metal dichalcogenides and oxides: a class, or individuals?

Superconductivity has “emerged” in several layered transition metal oxides and dichalcogenides, including new members in the last decade or so. The unusual and perhaps unique, single band triangular lattice system, Li_xNbO_2 is discussed immediately below. This one, along with several others that have unusual characteristics, has $T_c \sim 5$ K as shown in Fig. 5. This range of T_c is not impressive in itself, but the observation that superconductivity continues to pop up in strongly 2D TM oxides and chalcogenides where correlation effects are moderate to strong, suggests new physics. The systems we briefly discuss now are shown in Fig. 5 versus date of discovery. These materials do not seem to be very strongly connected to the cuprates (perhaps not at all), where T_c is a factor of 20–25 greater. But there are several examples. In the dichalcogenides, superconductivity arises in the same sort of systems, if not the same systems, where charge density waves (CDWs) and spin density waves (SDWs) are observed. These systems – at least the SDW members – display ordering wavevectors that are connected with Fermi surface calipers, and therefore are considered as Fermi surface instabilities. However, they are instabilities at q away from $\vec{q} = 0$, whereas superconductivity is a $\vec{q} = \vec{k} - \vec{k}' = 0$ instability (because pairing couples \vec{k} with $\vec{k}' = -\vec{k}$) since no translational symmetry is broken. Although superconductivity with pairing wavevector q different from zero is discussed more and more, this exotic FFLO (Fulde-Farrell-Larkin-Ovchinnikov [69, 70]) type of pairing is yet to be established in any system. The main idea behind FFLO pairing, for a system with spin imbalance and thus somewhat different up- and down-FSs, is that a loss of kinetic energy by forming pairs with non-zero center of mass can be compensated by retaining partial “nesting” of electron and hole FSs.

6.1 Li_xNbO_2 : a triangular lattice, single band correlated superconductor

The discovery of HTS in the cuprates in 1986 enlivened interest not only in layered cuprates but also in layered transition metal oxides more generally. The cuprates provided many-body theorists with a palette to study strong correlation effects in doped 2D antiferromagnets within

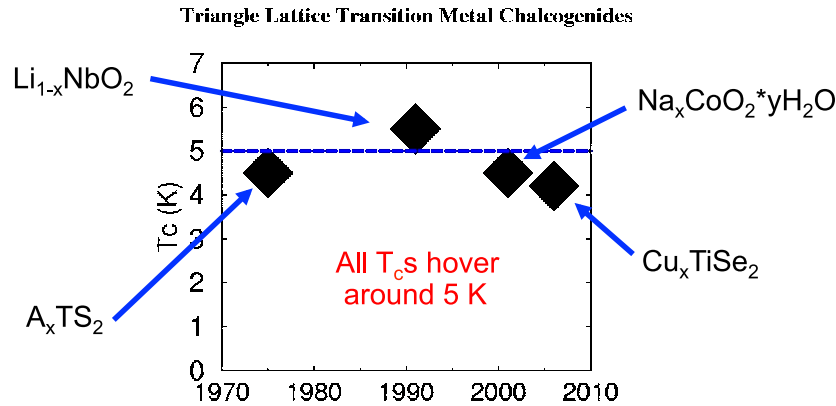
Synopsis: T_c in 2D Triangular Oxides/Chalcogenides

Fig. 5: Display of four classes of layered transition metal dioxides or dichalcogenides, showing T_c versus year of discovery. Each has its specific peculiarities: the dichalcogenides also host spin- and charge-density waves; the niobate is a unique single band triangular lattice system; the cobaltate must be hydrated to become superconducting. Their critical temperatures are all in the neighborhood of 5 K.

a single band model. The intricate physics that occurs in the low carrier density regime has come at the cost of a more direct effort to focus on identifying the pairing mechanism. (In more recent times multiband models have become more popular for the cuprates.)

Another favorite of many-body modelers is the triangular lattice, because with antiferromagnetic coupling magnetic order (in addition to simple charge and orbital order) is frustrated, and the observed or calculated phenomena become very rich. However, true single band systems are sparse, and finding one on a triangular lattice is rare indeed.

In 1990 Geselbracht *et al.* [71, 72] reported superconductivity up to 5.5 K in the Li_xNbO_2 system, synthesized and characterized structurally earlier by Meyer and Hoppe [73]. This system has been found [74] to be a single Nb $4d$ -band, triangular lattice system, which promises to display correlated electron behavior [75] because the calculated bandwidth is smaller than the anticipated intraatomic repulsion U on Nb. At stoichiometry, LiNbO_2 should be a d^2 low-spin (i.e. nonmagnetic) ionic band insulator. With all Li removed ($x = 0$ at fixed layered structure, see below), NbO_2 would be a d^1 compound and an excellent candidate as a Mott insulator; however, a rutile-related crystal structure is energetically favored in this limit. At intermediate concentrations it should conduct, unless charge order or some other exotic phase arises at certain band fillings.

This is, so far, a single member class – a unique example; cuprates after all have several subclasses and dozens of members – which unfortunately has seen little further experimental study [76] but a fair amount of theoretical investigation [74, 75, 77–79]. The charge carriers hop amongst the Nb sites, which form a triangular lattice such that electron dispersion is strongly two dimensional [74]. The unique aspect is that the triangular prismatic coordination creates a strong crystal field that leaves the $4d_{z^2}$ orbital lowest in energy and well separated from the other $4d$ bands above and the O $2p$ bands below. It becomes a single band, triangular lattice

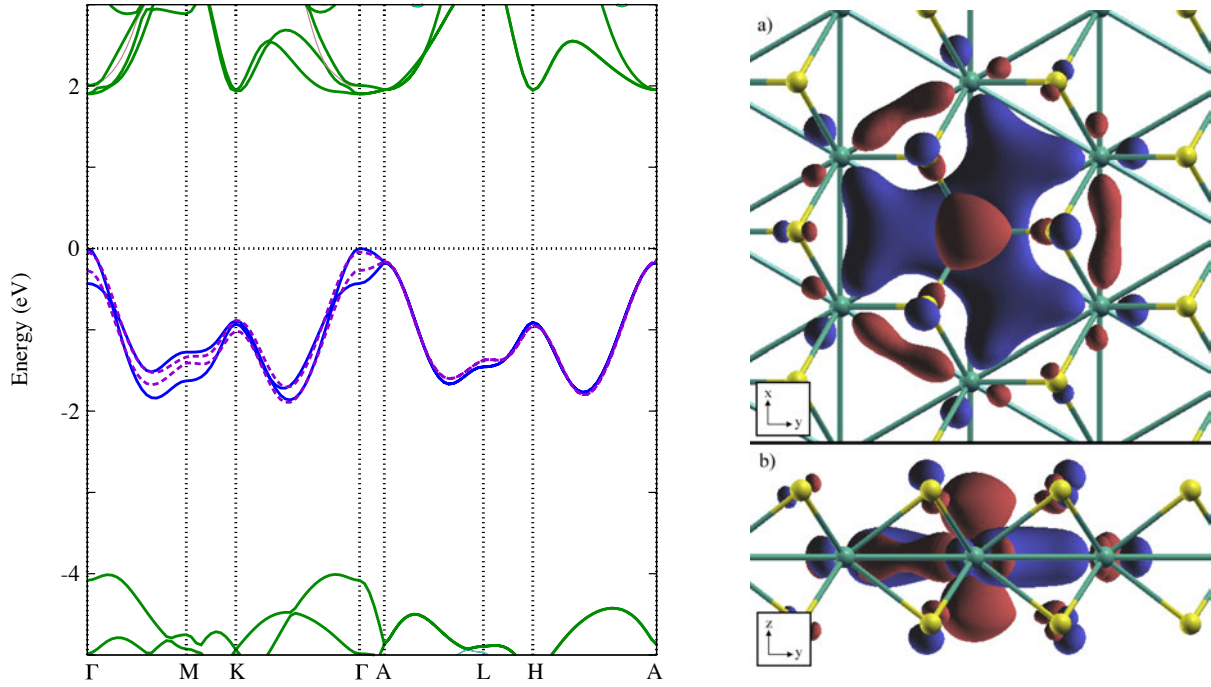


Fig. 6: Left panel: Band structure of LiNbO_2 for the experimental structure. The two central Nb d_{z^2} bands arise from the two Nb atoms in the unit cell, and lie within a 6 eV gap separating the valence O 2p bands from the other Nb 4d bands above. The result is a triangular lattice single band system. Right panel: Isosurface plot of the the d_{z^2} -symmetry Wannier function. The top subpanel provides a top view, revealing the large “fan blades” extending toward neighboring Nb ions, represented by small aqua-colored spheres. The bottom subpanel shows the d_{z^2} lobe projecting perpendicular to the Nb layers, and small contributions from neighboring O ions. Red and blue indicate opposite signs of the Wannier function. The Wannier function as a whole has “s-like” (fully symmetric) symmetry of the Nb site.

system with formal charge $\text{Nb}^{(4-x)+} : d^{1+x}$. As mentioned above, the $x = 0$ limit, which is not reached experimentally, corresponds to a triangular lattice Mott insulator according to the anticipated parameters [74] for this system: Hubbard U of 3–4 eV, DFT bandwidth of 1 eV. The observed T_c up to 5.5 K is reported to be insensitive to the band filling, according to the (somewhat sparse) data.

Being a very light element, Li is almost invisible to X-rays and, when samples are not of ideal quality (as these are not), Li concentration must be determined by other means. The multiphase nature of samples results in further uncertainty in the Li content of a given phase. Two other methods of doping the NbO_2 layer have been reported. One is that H is introduced into LiNbO_2 . The same $T_c = 5.5$ K results, and the supposition by Kumada *et al.* [80] is that this procedure also produces hole-doping (from the stoichiometric $x = 1$ compound), presumed to be due to formation of H^- . This supposition needs confirmation. In addition, $\text{Mg}_{0.5}\text{NbO}_2$ has been synthesized; this compound is structurally “identical” [81] to isovalent LiNbO_2 . A sharp negative swing in the susceptibility occurred at 4.4 K, but the authors declined to interpret this necessarily as superconductivity (although a small volume fraction of superconductivity seems to be another possible source). It is intriguing to note that $\text{Mg}_{0.5}\text{NbO}_2 \equiv \text{MgNb}_2\text{O}_4$ has one more

electron per transition metal than LiV_2O_4 , which is one of the very few $3d$ based heavy fermion compounds. The theoretical studies strongly suggest that the conducting phases of Li_xNbO_2 should be rather strongly correlated.

In 2009 a startling development in this system was announced. Xue *et al.* [82] reported values of T_c in the 14-17 K range, three times larger than earlier reports. Purity of the samples was sufficient to rule out $\text{NbC}_{1-y}\text{N}_y$, which has T_c in this same range, as the origin of the superconductivity. Moreover, the volume fraction of superconductivity was sufficient also to rule out the carbonitride phase. If confirmed (data always need independent confirmation) this higher range of T_c makes Li_xNbO_2 a much more interesting and important case.

6.2 Na_xCoO_2

This celebrated and heavily studied system is frustrating, in both senses of the word. As with the other triangular lattice compounds covered in this section, the transition metal (Co) sublattice [83] is frustrated for AFM coupling: the simplest way to see this is to note that around a triangle spin-ordering cannot proceed up-down-up-down because the 1st and 4th sites are the same. This fact, and extensions that arise from it, form the core of much of the interest in triangular lattice systems. Na_xCoO_2 is doubly frustrating for those hoping to understand its behavior because the superconductivity itself continues to present awkward aspects. Two criteria are necessary for superconductivity to appear: (i) the Na concentration must be near (usually somewhat larger than) $x \sim 1/3$, and (ii) the sample must be hydrated (i.e. dropped in water, or otherwise exposed to a great deal of water vapor H_2O). The observation that $x = 1/3$ might be special is supported by correlated band theory studies [84] that show a strong tendency in such a system, if strongly correlated, toward $\sqrt{3} \times \sqrt{3}$ charge and/or spin ordering (and perhaps orbital ordering). Ordering is also predicted at half filling, and indeed an ordered, Mott insulating phase is observed at $x = 1/2$. The frustrating thing is that it remains mysterious what incorporation of H_2O does – beyond the expectation that the molecule decomposes – still it is essential for superconductivity.

It is not in the purview of this lecture to survey the extensive experimental work on this system, nor the also rather extensive theoretical work. We do however point out that several experimental studies have tried to ascertain the *oxidation state* of the Co ion, versus the “doping level” x of Na. All have concluded that the oxidation state of Co is characteristic of a doping level (Na concentration plus things that H_2O might cause) of $x_{\text{eff}} \sim 0.55 - 60$, that is, moderately electron-doped above half filling of the relevant Co $3d$ band. This system remains a conundrum, one for which there are few if any solid models.

6.3 Doped transition metal dichalcogenides; recently Cu_xTiSe_2

This transition metal dichalcogenide class of quasi-2D materials, mostly metals, has a long history and large literature. Many examples of CDW and SDW materials occur in this system, and a glimpse of the many phenomena that occur in this system can be obtained from a recent report

on $1T\text{-Ta}_{1-x}\text{Fe}_x\text{S}_2$ [85], which contains a normal metal phase at high temperature and charge-ordered, superconducting, and correlated insulator phases at lower temperature. (The “1T,” “2H,” etc. designations indicate symmetry and stacking of the TaS_2 motifs.) SDW materials usually have a magnetic-order wavevector that can be identified with a Fermi surface caliper. The same had been suggested for CDW phases early on, and presumed for many years, since the generalized (Lindhard) susceptibility is expected to peak at wavevectors spanning the Fermi surface. This viewpoint has been questioned in recent years, and the complexity of the phase diagrams in dichalcogenides rivals those of oxides. Calculations of the susceptibility, including the relevant matrix elements, seem in several cases not to bear out earlier expectations: CDW wavevectors and Fermi surface calipers sometimes do not match up [86]. Recent evidence indicates that states not only near the Fermi surface but also some distance away (on an eV energy scale) contribute almost as heavily.

The electron-doped system Cu_xTiSe_2 system has caused some attention to return to this class of materials. The doping by Cu is proposed to allow study of the relevant phase diagram via simple (synthetic) means [87]; however, the solubility limit is only 11% for this compound. The superconducting T_c peaks in this system just above 4 K. The many questions greatly outweigh the few answers. One important experimental result is that the superconductivity is reported to be *s*-wave [88]; given the plethora of indications that electronic correlation effects are strong in this system, this “conventional” form of gap should not be interpreted as a strong indicator of electron-phonon pairing.

7 NaAlSi: unusual self-doped semimetallic superconductor

Occasionally a semimetal is encountered that is self-doped: a semimetal arising from “accidentally” overlapping bonding valence and antibonding conduction bands. It was noted in Sec. 4 that MgB_2 can also be regarded as a self-doped semimetal. Much more occasionally such a material is a superconductor; elemental Bi with its distorted fcc lattice is a well known, though not understood, example. NaAlSi which superconducts at 7 K [89], interestingly possesses the crystal structure of the “111” iron pnictide superconductors although their electronic structures have nothing in common. Another intriguing, but surely irrelevant, aspect of this compound is that moving each element to the next higher row (smaller Z , but isoelectronic) gives LiBC, the MgB_2 -like material that was discussed briefly earlier in this lecture.

Structurally, the AlSi_4 tetrahedra replace the FeAs_4 tetrahedra that form the basic feature of the “111” materials, while the buckled layer of interstitial Na ions simply contributes its electrons to the Al-Si bands. This provides 8 valence electrons per formula unit, which encourages covalent bonding and the formation of bonding valence and antibonding conduction bands. This indeed occurs although a simple characterization of the bonding-antibonding distinction has not yet been constructed. It is established that the bonding bands are strongly Si in character while the conducting bands are primarily Al. The gap is small, however, and the bands overlap slightly [90] near $\vec{k} = 0$, giving a semimetallic band structure.

The resulting density of states, shown in Fig. 7, is predicted from DFT studies to display an

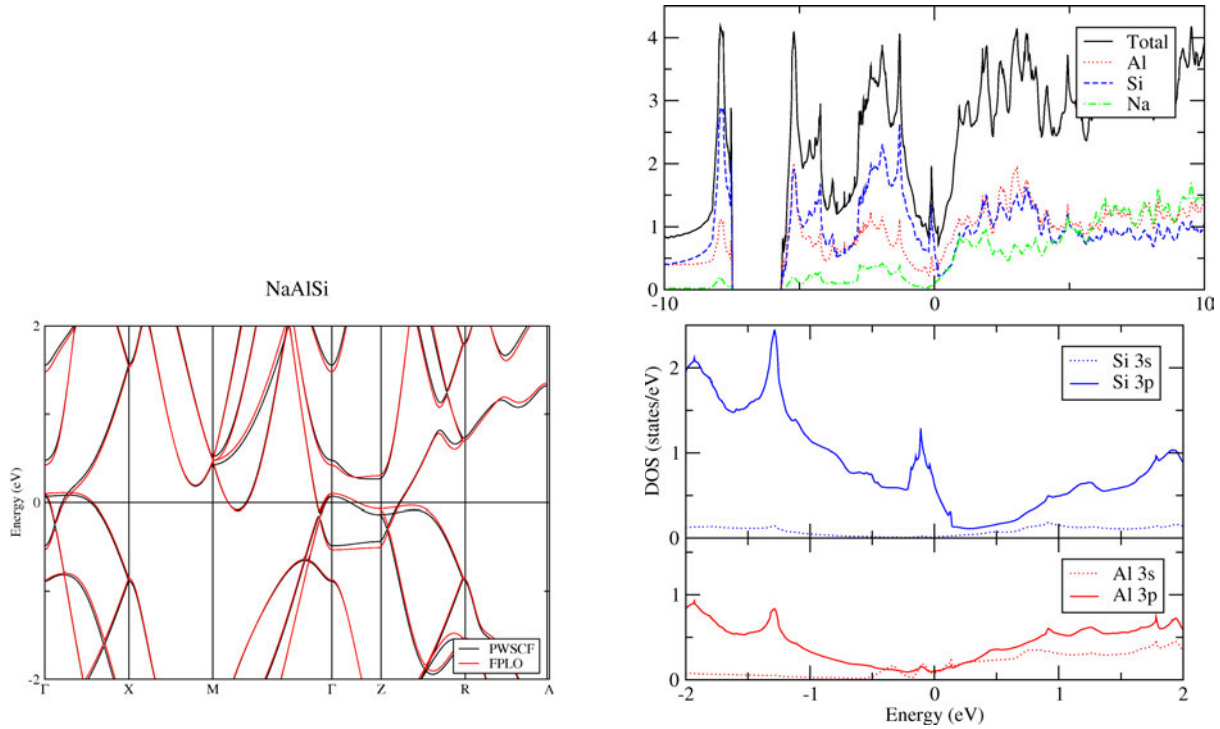


Fig. 7: *Left panel: Density functional based band structure of NaAlSi near the Fermi level. The band structure was calculated with the two methods (all-electron, and pseudopotential) that are designated in the caption. The conduction bands at and above the Fermi level are strongly Al in character, while the bands extending below the Fermi level are Si-derived. The energy scale is in eV. Right panel: Pictured in a 20 eV wide region, the total and atom- and orbital-projected density of states of NaAlSi. Note the pseudogap around the Fermi energy, with the sharp peak at the minimum. The middle subpanel provides an expanded view of the very narrow and sharp peak spanning the Fermi energy. The lower subpanel shows no contribution of Al states to the density of states peak.*

extremely sharp and narrow peak overlapping the Fermi level [90]. The scale of strong variation of $N(E)$ is similar to that of the largest phonon frequency, implying that the Born-Oppenheimer approximation underlying electron-phonon- and hence Eliashberg theory cannot necessarily be relied on. The superconductivity in NaAlSi requires further developments in theory. An attempt to evaluate the electron-phonon coupling strength using conventional theory (including the Born-Oppenheimer approximation) was thwarted by the small Fermi surfaces, which require finer \vec{k} and \vec{Q} meshes than were possible with even rather large computer clusters and memories. Another conundrum is presented by this system. The isostructural and isovalent sister compound NaAlGe has also been synthesized. Its electronic structure is virtually identical to that of NaAlSi. Nonetheless, it is found not to be superconducting (above 2 K). This fact revives the question occurring in the HfNCl and ZrNCl system: can the difference in superconducting behavior arise from the small and seemingly negligible differences in the electronic structure, or is it due to the mass difference – in this case Ge (73 amu) versus Si (28 amu), or to some other as yet undetermined origin? Another point of interest that we mention in passing is the relation, or perhaps not, to its relative, CaAlSi [91, 92], that has one more valence electron.

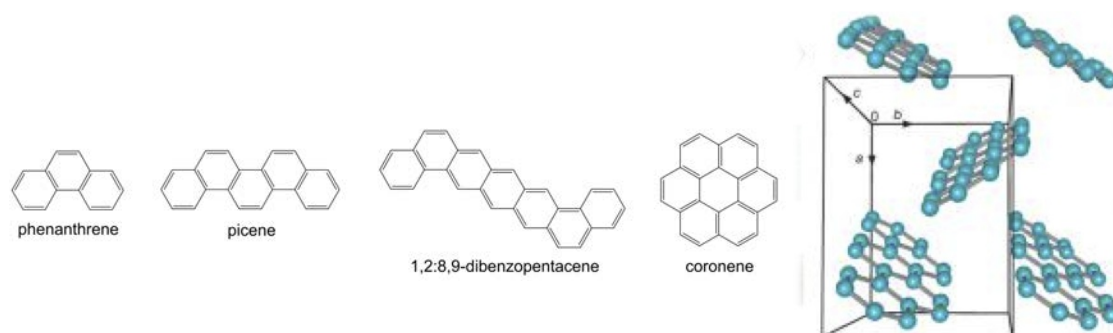


Fig. 8: *Left panel: Structures of the four hydrocarbon molecules that, when condensed to crystalline form and electron-doped with alkali atoms, superconduct. The coronene closed wheel of benzene rings is structurally distinct from the other “benzene chains.” Right panel: Crystal structure of picene, showing the herringbone alignment of molecules; only carbon atoms are pictured in both panels. The box outlines a primitive cell. The non-intuitive orientation and alignment of molecules results in the low symmetry monoclinic $P2_1$ space group.*

8 Doped hydrocarbons: organic crystals

A recent development, coming after the data used for Fig. 1 was available and which is only now beginning to create a stir, is the demonstration of T_c up to 33 K in electron-doped hydrocarbon solids. Superconductivity in organometallic compounds has been under study for well over two decades, and the originally low values of T_c had been raised to the 10 K regime. For reviews see the book by Ishiguro, Yamaji, and Saito [93] and the overview by Jerome [94]. These materials are strongly 2D in their electronic properties, and seem to show a combination of considerable correlated electron behavior as well as strong electron-phonon coupling. A coherent picture is lacking.

The recent developments center on molecular solids built of the aromatic hydrocarbon molecules phenanthrene $C_{14}H_{10}$, picene $C_{22}H_{14}$, and dibenzopentacene $C_{30}H_{18}$, comprising three, five, and seven connected benzene rings, respectively. For K_x picene, T_c up to 18 K was reported in 2010 by Mitsuhashi *et al.* [95], and this has been followed by Xue *et al.* in 2012 reporting $T_c = 33$ K in K_x dibenzopentacene [96]. These latter authors have noted that the maximum T_c so far appears to be linear in the number of benzene rings (each ring adding ~ 7 K) and they suggest that “delocalization” of the conduction electron wavefunctions over the molecule is a relevant factor. The molecules are sometimes described heuristically as tiny flakes or ribbons of H-capped graphene; however, they differ in containing C-C double bonds, see Fig. 8.

These systems, especially the picene-based one, are attracting active study from both experimentalists and theorists, and an overview is inadvisable at this time. It is relevant to this lecture, however, that density functional based linear response calculation of the phonon dispersion and electron-phonon interaction strength and spectral distribution have been reported by Subedi and Boeri [97]. They obtain strong coupling to H-C bend modes at 1400 cm^{-1} and C-C stretch modes around 1600 cm^{-1} and for various doping levels obtain coupling strength values in the range $\lambda \approx 0.65 - 0.75$, which is enough to account for the observed values of T_c . Whether these materials are really Fermi liquid metals (needed for the validity of Eliashberg theory) is currently being explored using several experimental techniques.

9 Summary of main points

From the data shown in Fig. 1, two dimensionality clearly seems to be special in producing classes of high temperature superconductors. Doped insulators account for a substantial number of these classes; the insulators may be either magnetic insulators (cuprates) or band insulators (TNCl). Beyond these two categories, the phenomena (and likely the pairing mechanisms) vary. The doped nitridochlorides do not display indications of the usual sort of strong electron correlation (enhancements; magnetic moments), while the doped insulators discussed in Sec. 6 fall within the categorization of electronically correlated materials. It is well recognized that the strongly correlated systems require more extensive study and that pairing mechanisms remain to be identified. One of the main purposes of this lecture is to point out that the transition metal nitridochlorides and similar materials are different, and seem to require their own distinct means of pairing.

Acknowledgments

For some unpublished results that have been included herein, I thank J.M. An (the FC_2 calculation), H. Rosner and A. Kitaigorodsky (MgB_2C_2), and E.R. Ylvisaker (BeB_2C_2). This overview is based partially upon a research project *Electron Pairing in Doped Ionic Insulators* funded by the U.S. National Science Foundation, which has followed after U.S. Department of Energy support through the SciDAC-e program. The author acknowledges support from the Simons Foundation during the preparation of this lecture.

References

- [1] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)
- [2] L.N. Cooper, Phys. Rev. **104**, 1189 (1956)
- [3] J. R. Schrieffer: *Superconductivity* (Benjamin, New York, 1964)
- [4] P.B. Allen, Phys. Rev. B **13**, 1416 (1976)
- [5] W. Kohn and J M. Luttinger, Phys. Rev. Lett. **15**, 524 (1965)
- [6] R. Joynt and L. Taillefer, Rev. Mod. Phys. **74**, 235 (2002)
- [7] G.M. Eliashberg, Zh. Eksperim. Teor. Fiz. **38**, 966 (1960)
English transl.: Soviet Phys. JETP **11**, 696 (1960)
- [8] D.J. Scalapino, J.R. Schrieffer, and J.W. Wilkins, Phys. Rev. **148**, 263 (1966)
- [9] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964)
- [10] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)
- [11] P.B. Allen and R.C. Dynes, Phys. Rev. B **12**, 905 (1975)
- [12] P.M. Platzman and T. Lenosky, Phys. Rev. B **52**, 10327 (1995)
- [13] Y. Takano, T. Takenouchi, S. Ishii, S. Ueda, T. Okutsu, I. Sakaguchi, H. Umezawa, H. Kawarada, and M. Tachiki, Diamond & Related Mat. **16**, 911 (2006)
- [14] W.E. Pickett, Physica C **468**, 126 (2008)
- [15] J.K. Jain and P.B. Allen, Phys. Rev. Lett. **54**, 2437 (1985)
- [16] P.B. Allen, M.L. Cohen, and D.R. Penn, Phys. Rev. B **38**, 2513 (1988)
- [17] A. Bill, H. Morawitz, and V. Kresin, Phys. Rev. B **66**, 100501 (2002)
- [18] A. Bill, H. Morawitz, and V. Kresin, Phys. Rev. B **68**, 144519 (2003)
- [19] I.N. Askerzade and B. Tanatar, Physica C **420**, 11 (2005)
- [20] C. Falter and G. A. Hoffmann, Phys. Rev. B **64**, 054516 (2001)
- [21] C. Falter, Phys. Stat. Sol. (b) **242**, 78 (2005)
- [22] G.S. Atwal and N W. Ashcroft, Phys. Rev. B **70**, 104513 (2004)
- [23] J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, and J. Akimitsu, Nature **410**, 63 (2001)

- [24] J.M. An and W.E. Pickett, Phys. Rev. Lett. **86**, 4366 (2001)
- [25] J. Kortus, I.I. Mazin, K.D. Belashchenko, V.P. Antropov, and L.L. Boyer, Phys. Rev. Lett. **86**, 4656 (2001)
- [26] K.D. Belashchenko, M. van Schilfgaarde, and V.P. Antropov, Phys. Rev. B **64**, 092503 (2001)
- [27] Y. Kong, O.V. Dolgov, O. Jepsen, and O.K. Andersen, Phys. Rev. B **64**, 020501 (2001)
- [28] A. Liu, I. I. Mazin, and J. Kortus, Phys. Rev. Lett. **87**, 087005 (2001)
- [29] W.E. Pickett, J. Supercond. & Novel Magn. **19**, 291 (2006)
- [30] M. Wörle, R. Nesper, G. Mair, M. Schwarz, and H.G. von Schnering, Z. Anorg. Allg. Chem. **621**, 1153 (1995)
- [31] H. Rosner, A. Kitaigorodsky, and W. E. Pickett, Phys. Rev. Lett. **88**, 127001 (2002)
- [32] J.M. An, S.Y. Savrasov, H. Rosner, and W.E. Pickett, Phys. Rev. B **66**, 220502 (2002)
- [33] A.M. Fogg, G.R. Darling, J.B. Claridge, J. Meldrum, and M.J. Rosseinsky, Phil Trans. Roy. Soc. A **366**, 55 (2008)
- [34] A. Lazicki, C.S. Yoo, H. Cynn, W.J. Evans, W.E. Pickett, J. Olamit, K. Liu, and Y. Ohishi, Phys. Rev. B **75**, 054507 (2007)
- [35] J. An and W.E. Pickett, unpublished.
- [36] M. Wörle and R. Nesper, J. Alloys Compounds **216**, 75 (1994)
- [37] P. Ravindran, P. Vajeeston, R. Vidya, A. Kjekshus, and H. Fjellvag, Phys. Rev. B **64**, 224509 (2001)
- [38] H. Harima, Physica C **378-381**, 18 (2002)
- [39] A.K. Verma, P. Modak, D.M. Gaitonde, R.S. Rao, B.K. Godwal, and L.C. Gupta, EPL **63**, 743 (2003)
- [40] E. Spano, M. Bernasconi, and E. Kopnin, Phys. Rev. B **72**, 014530 (2005)
- [41] S. Lebegue, B. Arnaud, and M. Alouani, Comp. Mat. Sci. **37**, 220 (2006)
- [42] T. Mori and E. Takayama-Muromachi, Current Appl. Phys. **4**, 276 (2004)
- [43] H. Y. Yan, M. G. Zhang, Q. Wei, and P. Guo, Comp. Mat. Sci. **68**, 174 (2013)
- [44] K. Hoffman, X. Rocquefelte, J.F. Halet, C. Bahtz, and B. Albert, Angew. Chemie - Intl. Ed. **47**, 2301 (2008)

- [45] A.H. Moudden, *Eur. Phys. J. B* **64**, 173 (2008)
- [46] E.R. Ylvisaker and W.E. Pickett, unpublished.
- [47] B. Albert and K. Schmitt, *Inorg. Chem.* **38**, 6159 (1999)
- [48] M. Somer, A. Yarasik, L. Akselrud, S. Leoni, H. Rosner, W. Schnelle, and R. Kniep, *Angew. Chem. - Int. Ed.* **43**, 1088 (2004)
- [49] S. Yamanaka, H. Kawaji, K. Hotehama and M. Ohashi, *Adv. Mater.* **9**, 771 (1996)
- [50] S. Shamoto, T. Kato, Y. Ono, Y. Miyazaki, K. Ohoyama, M. Ohashi, Y. Yamaguchi, and T. Kajitani, *Physica C* **306**, 7 (1998)
- [51] S. Yamanaka, K. Hotehama and H. Kawaji, *Nature* **392**, 580 (1998)
- [52] S. Shamoto, T. Kato, Y. Ono, Y. Miyazaki, K. Ohoyama, M. Ohashi, Y. Yamaguchi, and T. Kajitani, *J. Phys. Chem. Solids* **60**, 1431 (1998)
- [53] M. Ohashi, S. Yamanaka, and M. Hattori, *J. Ceram. Soc. Jpn. Int. Ed.* **97**, 1175 (1989)
- [54] S. Yamanaka, T. Yasunaga, K. Yamaguchi, and M. Tagawa, *J. Mater. Chem.* **19**, 2573 (2009)
- [55] A. Kaur, E. R. Ylsivaker, Y. Li, G. Galli, and W.E. Pickett, *Phys. Rev. B* **82**, 155125 (2010)
- [56] Y. Taguchi, A. Kitora, and Y. Isawa, *Phys. Rev. Lett.* **97**, 107001 (2006)
- [57] C.M. Schurz, L. Shlyk, T. Schleid, and R. Niewa, *Zeitschrift f. Krist.* **226**, 395 (2011)
- [58] R. Weht, A. Filippetti, and W.E. Pickett, *EPL* **48**, 320 (1999)
- [59] R. Heid and K.P. Bohnen, *Phys. Rev. B* **72**, 134527 (2005)
- [60] R. Akashi, K. Nakamura, R. Arita, and M. Imada, *Phys. Rev. B* **86**, 054513 (2012)
- [61] E.R. Ylvisaker and W.E. Pickett, *EPL* **101**, 57006 (2013)
- [62] Y. Taguchi, Y. Kasahara, T. Kishume, T. Takano, K. Kobayashi, E. Matsuoka, H. Onodera, K. Kuroki, and Y. Iwasa, *Physica C* **470**, S598 (2010)
- [63] Q. Yin, E.R. Ylvisaker, and W.E. Pickett, *Phys. Rev. B* **83**, 014509 (2011)
- [64] D.A. Headspith and M.G. Francesconi, *Topics Catal.* **52**, 1611 (2009)
- [65] T. Takasaki, T. Ekino, H. Fujii, and S. Yamanaka, *J. Phys. Soc. Japan* **74**, 2586 (2005)
- [66] Y. Taguchi, T. Kawabata, T. Takano, A. Kitora, K. Kato, M. Takata, and Y. Isawa, *Phys. Rev. B* **76**, 064508 (2007)

- [67] D.H. Gregory, M.G. Barker, P.P. Edwards, and D.J. Siddons, *Inorg. Chem.* **35**, 7608 (1996)
- [68] D.H. Gregory, P.M. O'Meara, A.G. Gordon, D.J. Siddons, A.J. Blake, M.G. Barker, T.A. Hamor, and P.P. Edwards, *J. Alloys Compds.* **317-318**, 237 (2001)
- [69] P. Fulde and R. A. Ferrell, *Phys. Rev.* **135**, A550 (1964)
- [70] A.I. Larkin and Yu.N. Ovchinnikov, *Zh. Eksp. Teor. Fiz.* **47**, 1136 (1964)
- [71] M.J. Geselbracht, T.J. Richardson, and A.M. Stacy, *Nature* **345**, 324 (1990)
- [72] M.J. Geselbracht, A.M. Stacy, A.R. Garcia, B.G. Silbernagel, and G.H. Kwei, *J. Phys. Chem.* **97**, 7102 (1993)
- [73] G. Meyer and R. Hoppe, *J. Less-Common Metals* **46**, 55 (1976)
- [74] E.R. Ylvisaker and W.E. Pickett, *Phys. Rev. B* **74**, 075104 (2006)
- [75] K.-W. Lee, J. Kuneš, R.T. Scalettar, and W.E. Pickett, *Phys. Rev. B* **76**, 144513 (2007)
- [76] E.G. Moshopoulou, P. Bordet, and J.J. Capponi, *Phys. Rev. B* **59**, 9590 (1999)
- [77] G.T. Liu, J.L. Luo, Z. Li, Y.Q. Guo, N.L. Wang, D. Jin, and T. Xiang, *Phys. Rev. B* **74**, 012504 (2006)
- [78] E.R. Ylvisaker, K.-W. Lee, and W.E. Pickett, *Physica B* **383**, 63 (2006)
- [79] D.L. Novikov, V.A. Gubanov, V.G. Zubkov, and A.J. Freeman, *Phys. Rev. B* **49**, 15830 (1994)
- [80] N. Kumada, S. Watauchi, I. Tanaka, and N. Kinomura, *Mater. Res. Bull.* **35**, 1743 (2000)
- [81] A. Miura, T. Takei, and N. Kumada, *J. Solid State Chem.* **197**, 471 (2013)
- [82] Z. Xue, A. Dong, Y. Guo, and G. Che, *J. Alloys Compds.* **476**, 519 (2009)
- [83] G. Meyer and R. Hoppe, *Angew. Chem., Int. Ed. Engl.* **13** (1974);
J. Less-Common Met. **46**, 55 (1976)
- [84] K.-W. Lee and W.E. Pickett, *Phys. Rev. Lett.* **96**, 096403 (2006)
- [85] R. Ang, Y. Tanaka, E. Ieki, K. Nakayama, T. Sato, L.J. Li, W.J. Lu, Y.P. Sun, and T. Takahashi, *Phys. Rev. Lett.* **109**, 176403 (2012)
- [86] M.D. Johannes and I.I. Mazin, *Phys. Rev. B* **77**, 165135 (2008)
- [87] E. Morosan, H.W. Zandbergen, B.S. Dennis, J.W.G. Bos, Y. Onose, T. Klimczuk, A.P. Ramirez, N.P. Ong, and R. J. Cava, *Nat. Phys.* **2**, 544 (2006)
- [88] S.Y. Li, G. Wu, X.H. Chen, and L. Taillefer, *Phys. Rev. Lett.* **99**, 107001 (2007)

- [89] S. Kuroiwa, H. Kawashima, H. Kinoshita, H. Okabe, and J. Akimitsu, *Physica C* **466**, 11 (2007)
- [90] H.B. Rhee, S. Banerjee, E.R. Ylvisaker, and W.E. Pickett, *Phys. Rev. B* **81**, 245114 (2010)
- [91] H. Sagayama, Y. Wakabayashi, H. Sawa, T. Kamiyama, A. Hoshikawa, S. Harjo, K. Uozato, A.K. Ghosh, M. Tokunaga, and T. Tamegai, *J. Phys. Soc. Jpn.* **75**, 043713 (2006)
- [92] S. Kuroiwa, H. Sagayama, T. Kakiuchi, H. Sawa, Y. Noda, and J. Akimitsu, *Phys. Rev. B* **74**, 014517 (2006)
- [93] T. Ishiguro, K. Yamaji, and G. Saito: *Organic Superconductors* (Springer-Verlag, Berlin, 1998)
- [94] D. Jérôme, *Chem. Rev.* **104**, 5565 (2004)
- [95] R. Mitsuhashi, Y. Suzuki, Y. Yamanari, H. Mitamura, T. Kambe, N. Ikeda, H. Okamoto, A. Fujiwara, M. Yamaji, N. Kawasaki, Y. Maniwa, and Y. Kubozono, *Nature* **464**, 76 (2010)
- [96] M. Xue, T. Cao, D. Wang, Y. Wu, H. Yang, X. Dong, J. He, F. Li, and G. F. Chen, *Sci. Rep.* **2**, 389 (2010)
- [97] A. Subedi and L. Boeri, *Phys. Rev. B* **84**, 020508 (2011)

12 Density Functional Perturbation Theory and Electron Phonon Coupling

Rolf Heid

Institute for Solid State Physics

Karlsruhe Institute of Technology

Contents

1	Introduction	2
1.1	Electron-ion Hamiltonian and adiabatic approximation	2
1.2	Phenomenological theory of lattice dynamics	4
2	Density functional perturbation theory	5
2.1	Lattice dynamics from first principles	5
2.2	Linear response formulation	8
2.3	Phonons in periodic lattices	10
3	Electron phonon coupling	12
3.1	Density functional perturbation approach to electron phonon vertex	12
3.2	Phonon self-energy and linewidth	15
3.3	Phonon mediated pairing interaction and superconductivity	18
3.4	Electron self-energy effects	22
4	Summary	27

1 Introduction

Electrons and ions are the fundamental building blocks of solids. The understanding of most solid state properties rests on the knowledge of the related quantum objects, electronic quasiparticles and phonons, respectively. Solving the quantum mechanical problem of the electron-ion coupling for extended systems is, however, a formidable task. Still, because of the large mass-difference between electrons and ions, they can be treated to a first approximation as independent dynamical subsystems. In the last decades, highly efficient numerical methods have been developed to solve the electronic part of the problem from first principles. Most of them are based on density functional theory and allow nowadays a routine investigation of the electronic structure of many compounds. The phonon problem took longer to be tackled from first principles, because an accurate solution of the electronic structure is a prerequisite for calculating the fundamental vibrational properties with sufficient accuracy. The development of a linear-response scheme, the so-called density functional perturbation theory, more than 20 years ago opened the door to efficient and accurate approaches and has matured into powerful numerical tools.

The interaction among these constituents, the electron-phonon coupling, influences or even dominates a variety of physical phenomena in solids. This is most noticeable in metals, where low-energy electronic excitations are strongly influenced by lattice vibrations – with important consequences for, e.g., electronic transport and thermodynamical properties. It also represents a natural source for electron pairing underlying the macroscopic quantum phenomenon of superconductivity.

In these lecture notes, I will give an introduction to the basic concepts underlying the modern numerical techniques to calculate phonons and electron-phonon coupling from first-principles within the framework of density functional theory. In Section 2, I will present an overview of the perturbational scheme to calculate phonon properties, and discuss some peculiarities of current implementations. Section 3 is devoted to the first principles approach to the electron-phonon coupling. Connection will be established to experimentally accessible quantities, like quasi-particle renormalization, and to the electron pairing interaction which enters the theory of superconductivity.

1.1 Electron-ion Hamiltonian and adiabatic approximation

We consider a solid to be build up from electrons and ions, where an ion consists of the nucleus and the tightly bound core electrons. The dynamics of electrons and ions in a crystal is described by the total Hamiltonian

$$\mathcal{H} = T_e + V_{ee} + T_i + V_{ii} + H_{e-i}, \quad (1)$$

where T_e and T_i are the kinetic energies of electrons and ions, respectively, V_{ee} denotes the Coulomb interaction among electrons, V_{ii} the interaction energy between ions, and H_{e-i} the interaction between electrons and ions.

The task of finding solutions for the Schrödinger equation $\mathcal{H}\Psi(\underline{\mathbf{r}}, \underline{\mathbf{R}}) = \mathcal{E}\Psi(\underline{\mathbf{r}}, \underline{\mathbf{R}})$, where $\underline{\mathbf{r}}$ and

$\underline{\mathbf{R}}$ stand for the set of electron and ion coordinates, respectively, can be drastically simplified due to the large difference of the electron mass m and the ion mass M : The light electrons can be considered as moving much faster than the heavy ions. They follow instantaneously the motion of the ions, while the latter perform small vibrations around their rest positions. As first shown by Born and Oppenheimer [1] for molecules and later applied to solids by Chester and Houghton [2], this picture can be proven by introducing a small parameter \varkappa which scales to 0 for $M \rightarrow \infty$. To this end, they considered small displacements of the ions from their rest positions of the form

$$\mathbf{R}_i = \mathbf{R}_i^0 + \varkappa \mathbf{u}_i. \quad (2)$$

A proper form for \varkappa can be inferred from the requirement, that the kinetic energy of the ions should be of the same order as the potential term: quadratic in \mathbf{u} . This leads to the choice $\varkappa = (m/M)^{1/4}$, which is less than 0.1 for all elements except H and He. One can now perform a systematic expansion of the Hamiltonian and wavefunctions in terms of this small parameter. To lowest order, the total wavefunction can be written as a product $\Psi(\mathbf{r}, \underline{\mathbf{R}}) = \chi(\underline{\mathbf{R}})\psi(\mathbf{r}; \underline{\mathbf{R}})$, where the electronic wavefunction depends only parametrically on the ion coordinates. The electronic wavefunction obeys the equation

$$[T_e + V_{ee} + H_{e-i}(\underline{\mathbf{R}})]\psi_n(\mathbf{r}; \underline{\mathbf{R}}) = E_n(\underline{\mathbf{R}})\psi_n(\mathbf{r}; \underline{\mathbf{R}}), \quad (3)$$

where the dependence on $\underline{\mathbf{R}}$ enters via the interaction H_{e-i} . The ion wavefunction is a solution of

$$[T_i + V_{ii} + E_n(\underline{\mathbf{R}})]\chi(\underline{\mathbf{R}}) = \mathcal{E}\chi(\underline{\mathbf{R}}). \quad (4)$$

This level of approximation is called the adiabatic or Born-Oppenheimer approximation. It describes a decoupling of the dynamics of the electrons and ions and neglects electronic excitations induced by the ionic motion. The electron system enters in (4) via the energies $E_n(\underline{\mathbf{R}})$ of the n -th eigenstate. Usually, one can resort to the ground state and drop the index n , because "normally" encountered excited states at finite temperatures do not deviate much on the scale relevant for the ionic motion. Nevertheless this term includes the important effect of screening of the ionic motion by the valence electrons, which is, however, the same for the ground state as for the excited states.

To go beyond the adiabatic approximation, one uses the solutions of (3) to expand the wavefunction of the solid in the form

$$\Psi_m(\mathbf{r}; \underline{\mathbf{R}}) = \sum_n \chi_{mn}(\underline{\mathbf{R}})\psi_n(\mathbf{r}; \underline{\mathbf{R}}). \quad (5)$$

The eigenvalue problem $\mathcal{H}\Psi_m = \mathcal{E}_m\Psi_m$ leads to the following equation for the ionic part

$$[T_i + V_{ii} + E_n(\underline{\mathbf{R}})]\chi_{mn}(\underline{\mathbf{R}}) + \sum_{n'} \Delta H_{nn'}\chi_{mn'}(\underline{\mathbf{R}}) = \mathcal{E}_m\chi_{mn}(\underline{\mathbf{R}}). \quad (6)$$

The new feature with respect to (4) is the appearance of two additional terms $\Delta H = \Delta H^{(1)} +$

$\Delta H^{(2)}$ given by

$$\Delta H_{nn'}^{(1)} = -\frac{1}{M} \sum_i \int d\mathbf{r}^{3N} \psi_n^*(\mathbf{r}; \mathbf{R}) \nabla_{R_i} \psi_{n'}(\mathbf{r}; \mathbf{R}) \cdot \nabla_{R_i} \quad (7)$$

$$\Delta H_{nn'}^{(2)} = -\frac{1}{2M} \sum_i \int d\mathbf{r}^{3N} \psi_n^*(\mathbf{r}; \mathbf{R}) \nabla_{R_i}^2 \psi_{n'}(\mathbf{r}; \mathbf{R}). \quad (8)$$

They contain derivatives of the electronic wavefunctions with respect to the ion coordinates, and take into account possible excitations in the electronic subsystem due to the motion of the ions. Among these two non-adiabatic terms $\Delta H^{(1)}$ is typically the dominant one, because from the expansion of ψ_n it contains terms of order κ , while $\Delta H^{(2)}$ involves terms of order κ^2 .

Corrections to $\Psi(\mathbf{r}, \mathbf{R})$ beyond the adiabatic approximation can be shown to be of order κ^3 and corrections to the energy are of order κ^6 . The expansion parameter κ only depends on the mass ratio and not on the strength of the electron-phonon interaction. Thus the adiabatic approximation is adequate for both free-electron-like systems and for compounds possessing tighter bound valence electrons like transition metals.

1.2 Phenomenological theory of lattice dynamics

Within the adiabatic approximation, the statics and dynamics of the ions are governed by an effective potential

$$\Omega(\mathbf{R}) = V_{ii}(\mathbf{R}) + E_0(\mathbf{R}), \quad (9)$$

where $E_0(\mathbf{R})$ denotes the electronic ground-state energy for a given ion configuration \mathbf{R} . The effective potential Ω builds the starting point of the microscopic theory of lattice dynamics, which has been outlined in a number of review articles [3–5].

Dynamical properties are derived by a systematic expansion of Ω for atom displacements \mathbf{u} around a chosen reference configuration, $\mathbf{R}_i = \mathbf{R}_i^0 + \mathbf{u}_i$, leading to

$$\Omega(\mathbf{R}) = \Omega(\mathbf{R}^0) + \sum_{i\alpha} \Phi_\alpha(i) u_{i\alpha} + \frac{1}{2} \sum_{i\alpha j\beta} \Phi_{\alpha\beta}(i, j) u_{i\alpha} u_{j\beta} + \dots \quad (10)$$

Greek indices α and β denote Cartesian coordinates, while i and j are atom indices. The term of first order is the negative of the force acting on an atom in the reference configuration

$$F_{i\alpha} = -\left. \frac{\partial \Omega}{\partial R_{i\alpha}} \right|_0 = -\Phi_\alpha(i). \quad (11)$$

It vanishes if one chooses as reference the equilibrium configuration that minimizes Ω . The second-order coefficients are given by

$$\Phi_{\alpha\beta}(i, j) = \left. \frac{\partial^2 \Omega}{\partial R_{i\alpha} \partial R_{j\beta}} \right|_0. \quad (12)$$

Their physical meaning becomes more evident when one considers the case where only a single ion at site i is displaced from the equilibrium position by $u_{i\alpha}$. Then the force felt by an atom at site j is given by:

$$F_{j\beta} = -\sum_{i\alpha} \Phi_{\alpha\beta}(i, j) u_{i\alpha}. \quad (13)$$

Thus, to lowest order, $\Phi_{\alpha\beta}(i, j)$ describes a linear relationship between displacement and resulting force. They are the 3D equivalent of a spring constant and are called harmonic force constants. Higher-order coefficients are denoted as anharmonic force constants. The harmonic approximation is based on truncating the sum after the second order.

In periodic crystals, the atoms are characterized by two indices $i = (l\kappa)$, which denote the unit cell (l) and the atoms inside a unit cell (κ), respectively. For periodic boundary conditions, the Fourier transform of the force constant matrix is related to the dynamical matrix

$$D_{\kappa\alpha\kappa'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_\kappa M_{\kappa'}}} \sum_l \Phi_{\alpha\beta}(l\kappa, 0\kappa') e^{-i\mathbf{q}(\mathbf{R}_{l\kappa}^0 - \mathbf{R}_{0\kappa'}^0)}, \quad (14)$$

which determines the equation for the normal modes or phonons,

$$\sum_{\kappa'\beta} D_{\kappa\alpha\kappa'\beta}(\mathbf{q}) \eta_{\kappa'\beta}(\mathbf{q}j) = \omega_{\mathbf{q}j}^2 \eta_{\kappa\alpha}(\mathbf{q}j). \quad (15)$$

$\omega_{\mathbf{q}j}$ and $\eta_{\kappa\alpha}(\mathbf{q}j)$ denote the energy and polarization of the normal mode determined by the wavevector \mathbf{q} and branch index j .

These quantities enter into the relationship between the atom displacements and the usual phonon annihilation and creation operators $b_{\mathbf{q}j}$ and $b_{\mathbf{q}j}^\dagger$, describing quantized normal modes

$$u_{l\kappa\alpha} = e^{i\mathbf{q}\mathbf{R}_{l\kappa}^0} \frac{1}{\sqrt{N_q}} \sum_{\mathbf{q}j} A_{\kappa\alpha}^{\mathbf{q}j} (b_{\mathbf{q}j} + b_{-\mathbf{q}j}^\dagger) \quad \text{with} \quad A_{\kappa\alpha}^{\mathbf{q}j} = \frac{\eta_{\kappa\alpha}(\mathbf{q}j)}{\sqrt{2M_\kappa\omega_{\mathbf{q}j}}}. \quad (16)$$

A complete characterization of the harmonic vibrational spectrum requires the knowledge of either the normal modes for the whole Brillouin zone, or the force constants for all atom bonds. For a metallic system, the latter representation is often more economical since the lattice interaction in real space is rather short ranged due to electronic screening.

2 Density functional perturbation theory

2.1 Lattice dynamics from first principles

The goal is now to calculate the basic quantities determining the dynamics of the ions. The first term in the effective potential (9) is the Coulomb interaction among the ions, whose contribution to the force constants can be readily obtained. The second term represents the electronic contribution, which incorporates all important physical properties like bonding and screening. It requires a sophisticated and accurate treatment of the electronic system, as provided by density functional theory.

2.1.1 Basics of density functional theory

The foundations of density functional theory (DFT) have been worked out by Hohenberg, Kohn, and Sham [6, 7] in the mid 60's, and are outlined in numerous reviews [8–10]. Here we only mention the essential features which we need later.

In DFT, the ground-state energy of a system of interacting electrons moving in an external potential v_{ext} is obtained by minimizing the functional

$$E[n] = F[n] + \int d^3r v_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) \quad (17)$$

with respect to the electron density $n(\mathbf{r})$. At its minimum, $n(\mathbf{r})$ is the true electron density of the interacting system. The functional $F[n]$ is universal, i.e. independent of the external potential. For practical applications, the scheme developed by Kohn and Sham has proven to be very useful. They showed that the minimum principle allows us to map the complex many-body problem onto a fictitious system of non-interacting electrons that in its ground state possesses the same inhomogeneous density as the interacting system [7]. They expressed the energy functional as

$$F[n] = T_s[n] + E_H[n] + E_{xc}[n], \quad (18)$$

where T_s represents the kinetic energy of the non-interacting electrons (we adopt Rydberg atomic units defined by $\hbar^2 = 2m = e^2/2 = 1$)

$$T_s[n] = \sum_i f_i \int d^3r \psi_i^*(\mathbf{r}) (-\nabla^2) \psi_i(\mathbf{r}) \quad (19)$$

and $E_H[n]$ the Hartree energy

$$E_H[n] = \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (20)$$

with the single-particle representation of the density

$$n(\mathbf{r}) = \sum_i f_i |\psi_i(\mathbf{r})|^2. \quad (21)$$

Here f_i denotes the occupation number of the single-particle state ψ_i . The wavefunctions of the fictitious electrons obey a single-particle equation (Kohn-Sham equation)

$$\left\{ -\nabla^2 + v_{\text{eff}}(\mathbf{r}) \right\} \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}). \quad (22)$$

The effective potential $v_{\text{eff}}(\mathbf{r})$ is a functional of the density and given as a sum of the external potential and a screening potential

$$v_{\text{eff}}[n] = v_{\text{ext}} + v_{\text{scr}}[n] = v_{\text{ext}} + v_H[n] + v_{XC}[n]. \quad (23)$$

The screening potential is obtained via functional derivatives of the last two terms in the total energy functional (18). It consists of the Hartree potential

$$v_H(\mathbf{r})[n] = \frac{\delta E_H}{\delta n(\mathbf{r})} = \int d^3r' \frac{2n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (24)$$

which describes an average electrostatic potential originating from the other electrons, and the exchange-correlation potential $v_{XC}(\mathbf{r}) = \delta E_{XC} / \delta n(\mathbf{r})$.

By this formulation, the original many-body problem has been cast into a set of single-particle equations (21)–(23) that has to be solved self-consistently. The complexity of the original many-body problem is transferred to the task of determining the exchange-correlation energy E_{XC} . The big success of DFT partly rests on the empirical fact that already simple approximations to v_{XC} often give very accurate results. The most widely used ansatz is the local-density approximation (LDA)

$$v_{XC}^{\text{LDA}}(\mathbf{r}) = \left. \frac{d(n\epsilon_{XC}^{\text{hom}}(n))}{dn} \right|_{n=n(\mathbf{r})}, \quad (25)$$

where $\epsilon_{XC}^{\text{hom}}(n)$ represents the exchange-correlation energy density of the homogeneous interacting electron gas. For $\epsilon_{XC}^{\text{hom}}$ various parametrizations derived from analytical and numerical studies exist [9]. Another popular ansatz is the generalized-gradient approximation (GGA), where in addition to LDA a dependence of v_{XC} on the local gradient of the electron density is considered to better account for inhomogeneous density distributions [11–13].

2.1.2 Application to lattice dynamics

As we have seen in Sec. 1.2, lattice-dynamical properties are determined by the adiabatic lattice potential Ω , which equals the ground-state energy for a fixed ion configuration. Hence, lattice dynamics depends only on ground-state properties of the electronic system, and is accessible in the framework of density functional theory. An overview of the various methods to extract lattice-dynamical properties from *ab-initio* calculations have been given in [14]. One can divide them into two main classes: (i) direct methods and (ii) linear-response techniques.

The direct methods are based on ground-state calculations for the ideal crystal and for geometries with ions displaced from their equilibrium position. The frozen-phonon (FP) technique is conceptually the simplest and historically the first-applied method and uses the quadratic dependence of the total energy from the displacement to extract the frequency of a normal mode [15]. Since this requires *a priori* knowledge of the phonon eigenvector, it can typically be used only in cases where symmetry completely determines its form.

A more efficient scheme employs the linear relationship (13) between an ionic displacement and the forces felt by other ions in the unit cell [16–18]. This can be achieved with little numerical expenses, as forces can be derived directly from quantities obtained in a ground-state calculation with the help of the Hellman-Feynman theorem. A single calculation then gives information about a complete row of the dynamical matrix. The complete dynamical matrix can be constructed using a few appropriately chosen displacements. Hence, this approach does not require any *a priori* information about the normal modes. Since frozen-phonon calculations employ finite displacements of ions, their results principally contain all anharmonic effects, which could be used to extract higher anharmonic coupling constants.

A disadvantage of the direct methods is the need to resort to supercells to extract properties for non-zero wavevector phonons. A complete determination of the phonon spectrum requires supercells with sizes larger than the effective range of the lattice interactions [19–22].

The alternative approach consists of calculating the derivatives of the total energy directly within perturbative schemes. In particular, the dynamical matrix is obtained from the second deriva-

tives via Eq. (12). It has the big advantage that it works directly in reciprocal space and gives access to the dynamical matrix at arbitrary wavevectors without the need for supercells. We will discuss this scheme in some detail now.

2.2 Linear response formulation

Here we show how the perturbative approach is set up within the DFT framework. We will first present some general considerations before applying them to the specific case of perturbations induced by ionic displacements in periodic crystals.

2.2.1 Energy derivatives

Let us consider a situation where the external potential v_{ext} depends on a set of adiabatic perturbation parameters $\Lambda = \{\lambda_a, a = 1, \dots, p\}$. Each v_{ext}^Λ determines an electronic ground state with density $n^\Lambda(\mathbf{r})$ and total energy $E^\Lambda = F[n^\Lambda] + \int d^3r n^\Lambda(\mathbf{r}) v_{\text{ext}}^\Lambda(\mathbf{r})$, which depends on the perturbation via the external potential and implicitly via the density. Its derivative then contains two contributions

$$\frac{\partial E^\Lambda}{\partial \lambda_a} = \int d^3r n^\Lambda(\mathbf{r}) \frac{\partial v_{\text{ext}}^\Lambda(\mathbf{r})}{\partial \lambda_a} + \int d^3r \frac{\delta E^\Lambda}{\delta n(\mathbf{r})} \frac{\partial n^\Lambda(\mathbf{r})}{\partial \lambda_a}. \quad (26)$$

Due to the variational principle, the second term vanishes for each finite Λ . Thus the first derivative depends on the ground-state density only. This represents the DFT equivalent of the well known Hellman-Feynman-Theorem [23].

The second-order derivatives are then given by

$$\frac{\partial^2 E^\Lambda}{\partial \lambda_a \partial \lambda_b} = \int d^3r \frac{\partial n^\Lambda(\mathbf{r})}{\partial \lambda_b} \frac{\partial v_{\text{ext}}^\Lambda(\mathbf{r})}{\partial \lambda_a} + \int d^3r n^\Lambda(\mathbf{r}) \frac{\partial^2 v_{\text{ext}}^\Lambda(\mathbf{r})}{\partial \lambda_a \partial \lambda_b}. \quad (27)$$

For practical purposes it is important that the second derivatives require only the knowledge of the first-order variations of the electron density. Therefore, it is sufficient to consider only the linear response of the electron system.

2.2.2 Linear response within the Kohn-Sham scheme

The linear response within the DFT scheme is obtained by standard perturbation techniques under the condition that the effective potential entering the Kohn-Sham equations depends on the ground-state density itself. Thus its linear variation is given by

$$\begin{aligned} \delta v_{\text{eff}}(\mathbf{r}) &= \delta v_{\text{ext}}(\mathbf{r}) + \delta v_{\text{scr}}(\mathbf{r}) = \delta v_{\text{ext}}(\mathbf{r}) + \int d^3r' I(\mathbf{r}, \mathbf{r}') \delta n(\mathbf{r}') \\ I(\mathbf{r}, \mathbf{r}') &\equiv \frac{\delta v_{\text{scr}}(\mathbf{r})}{\delta n(\mathbf{r}')} = \frac{\delta v_{\text{H}}(\mathbf{r})}{\delta n(\mathbf{r}')} + \frac{\delta v_{\text{XC}}(\mathbf{r})}{\delta n(\mathbf{r}')} = \frac{2}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta^2 E_{\text{XC}}}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')}. \end{aligned} \quad (28)$$

This induces a first-order variation of the single-particle wavefunctions

$$\delta \psi_i(\mathbf{r}) = \sum_{j(\neq i)} \frac{\langle j | \delta v_{\text{eff}} | i \rangle}{\epsilon_i - \epsilon_j} \psi_j(\mathbf{r}). \quad (29)$$

Using a similar expression for $\delta\psi_i^*(\mathbf{r})$ gives

$$\begin{aligned}\delta n(\mathbf{r}) &= \sum_i f_i [\psi_i^*(\mathbf{r}) \delta\psi_i(\mathbf{r}) + \delta\psi_i^*(\mathbf{r}) \psi_i(\mathbf{r})] \\ &= \sum_{i \neq j} \frac{f_i - f_j}{\epsilon_i - \epsilon_j} \langle j | \delta v_{\text{eff}} | i \rangle \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}).\end{aligned}\quad (30)$$

Eqs. (28) and (30) must be solved self-consistently to obtain the first-order variation of the density. To proceed, one can write the linear relationship (30) between δn and δv_{eff} more explicitly

$$\delta n(\mathbf{r}) = \int d^3r' \chi_0(\mathbf{r}, \mathbf{r}') \delta v_{\text{eff}}(\mathbf{r}') \quad (31)$$

$$\chi_0(\mathbf{r}, \mathbf{r}') = \sum_{i \neq j} \frac{f_i - f_j}{\epsilon_i - \epsilon_j} \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}'). \quad (32)$$

Here, χ_0 represents the charge susceptibility of the non-interacting Kohn-Sham system. It is expressed solely by ground-state quantities [24]. In the case of a periodic system, this is just the well-known Adler-Wiser form [25, 26]. Although obtained by perturbation theory, Eq. (32) is exact because the Kohn-Sham equations describe non-interacting electrons.

In combination with Eq. (28) this leads to

$$\delta v_{\text{eff}} = \delta v_{\text{ext}} + I \chi_0 \delta v_{\text{eff}}, \quad (33)$$

which can be solved for δv_{eff}

$$\delta v_{\text{eff}} = [1 - I \chi_0]^{-1} \delta v_{\text{ext}} = \epsilon^{-1} \delta v_{\text{ext}}, \quad (34)$$

where $\epsilon = 1 - I \chi_0$ denotes the static dielectric matrix and describes the screening of the "bare" perturbation from the external potential.

The problem is now reduced to a calculation of ϵ^{-1} . Historically this was the first route to be explored [27, 28]. Direct application of these equations, however, has several practical disadvantages. It requires an inversion of the matrix $\epsilon(\mathbf{r}, \mathbf{r}')$, which for periodic systems is most conveniently done in Fourier space. This inversion turns out to be the bottleneck of this scheme, as a proper convergence often requires a large number of Fourier components. Attempts to perform this inversion in direct space using a Wannier representation did not lead to significant improvements [29]. In the calculation of χ_0 in Eq. (32), unoccupied orbitals do enter, which are not available for bandstructure methods employing minimal basis sets (e.g. LMTO).

2.2.3 Modern formulation: Density functional perturbation theory

An important progress has been achieved by a new formulation of the linear-response approach that avoids some of the aforementioned problems of the dielectric matrix approach. It is called density functional perturbation theory (DFPT) and has been proposed independently by Zein *et al.* [30–32] and Baroni *et al.* [33, 34]. A concise description can be found in [35]. We will give a short outline for the case of a non-metallic system.

The expression (30) for the first-order density variation contains a double sum over electronic states. The prefactor $(f_i - f_j)/(\epsilon_i - \epsilon_j)$ restricts it to combinations where one state comes from the valence space and the other from the conduction space. Using time-reversal symmetry, this can be rewritten as

$$\delta n(\mathbf{r}) = 2 \sum_{vc} \frac{1}{\epsilon_v - \epsilon_c} \langle c | \delta v_{\text{eff}} | v \rangle \psi_v^*(\mathbf{r}) \psi_c(\mathbf{r}). \quad (35)$$

To avoid summation over the conduction states, one rewrites

$$\delta n(\mathbf{r}) = 2 \sum_v \psi_v^*(\mathbf{r}) \Delta_v(\mathbf{r}) \quad (36)$$

with

$$|\Delta_v\rangle = \sum_c \frac{1}{\epsilon_v - \epsilon_c} |c\rangle \langle c | \delta v_{\text{eff}} | v \rangle. \quad (37)$$

This quantity fulfills the following linear equation:

$$(H - \epsilon_v) |\Delta_v\rangle = - \sum_c |c\rangle \langle c | \delta v_{\text{eff}} | v \rangle = -P_c \delta v_{\text{eff}} | v \rangle = (P_v - 1) \delta v_{\text{eff}} | v \rangle. \quad (38)$$

Here $P_c = \sum_c |c\rangle \langle c|$ denotes the projector onto the conduction space, and $P_v = 1 - P_c$ is the projector onto the valence space. By this reformulation, only valence-state quantities enter the equation for Δ_v , and one avoids an expensive summation over conduction states.

2.2.4 Beyond linear response: $(2n + 1)$ theorem

As shown above, the first derivative of the energy depends solely on the unperturbed ground-state density, while second-order derivatives require knowledge of the density and its first-order derivatives. Both results are special cases of the so-called $(2n + 1)$ theorem, which states that all derivatives of the total energy up to $(2n + 1)$ -th order with respect to the adiabatic perturbation can be calculated from the knowledge of all derivatives of the Kohn-Sham eigenstates and density up to n -th order. In the framework of density-functional theory this theorem also holds for nonlocal external potentials and is thus applicable within pseudopotential methods. The proof given by Gonze *et al.* [36–38] essentially rests on the variational property of the energy functional.

As a corollary of this theorem, harmonic as well as third-order anharmonic force constants merely require calculation of the linear variations of the Kohn-Sham eigenstates and the density. Both are accessible by linear-response calculations.

2.3 Phonons in periodic lattices

Here we discuss the details of the calculations of the interatomic force constants within the density functional perturbation approach. To this end, we consider periodic displacements of the ions from their equilibrium positions, $\mathbf{R}_{l\kappa} = \mathbf{R}_{l\kappa}^0 + \mathbf{u}_{l\kappa}$, of the form

$$u_{l\kappa\alpha} = d_{\kappa\alpha} e^{i\mathbf{q}\mathbf{R}_{l\kappa}^0} + d_{\kappa\alpha}^* e^{-i\mathbf{q}\mathbf{R}_{l\kappa}^0}, \quad (39)$$

where l denotes the unit cell, κ specifies the ion inside a unit cell, and α indicates Cartesian coordinates. The complex amplitudes $d_{\kappa\alpha}$ allow us to vary the relative phase of the displacement. It is convenient to denote the corresponding derivatives by $\delta_{\kappa\alpha}^{\mathbf{q}} \equiv \frac{\partial}{\partial d_{\kappa\alpha}}$ and $\delta_{\kappa'\beta}^{-\mathbf{q}} \equiv \frac{\partial}{\partial d_{\kappa'\beta}^*}$. The electronic contribution to the dynamical matrix can be then written as a mixed derivative

$$D_{\kappa\alpha\kappa'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_{\kappa}M_{\kappa'}}} \delta_{\kappa\alpha}^{\mathbf{q}} \delta_{\kappa'\beta}^{-\mathbf{q}} E \Big|_{\mathbf{u}=0}. \quad (40)$$

Usually, the external potential is expressed as a superposition of atomic potentials v_{κ} centered at the instantaneous positions of the ions

$$v_{\text{ext}}(\mathbf{r}) = \sum_{l\kappa} v_{\kappa}(\mathbf{r} - \mathbf{R}_{l\kappa}). \quad (41)$$

Its first-order variation, evaluated at the equilibrium positions, is given by

$$\begin{aligned} \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{ext}}(\mathbf{r}) &= - \sum_l \nabla_{\alpha}^{\mathbf{r}} v_{\kappa}(\mathbf{r} - \mathbf{R}_{l\kappa}^0) e^{i\mathbf{q}\mathbf{R}_{l\kappa}^0} \\ &= -e^{i\mathbf{q}\mathbf{r}} \sum_l e^{i\mathbf{q}(\mathbf{R}_{l\kappa}^0 - \mathbf{r})} \nabla_{\alpha}^{\mathbf{r}} v_{\kappa}(\mathbf{r} - \mathbf{R}_{l\kappa}^0). \end{aligned} \quad (42)$$

The quantity defined by the lattice sum has the periodicity of the original lattice. Thus the derivative $\delta_{\kappa\alpha}^{\mathbf{q}}$ can be considered to carry a momentum \mathbf{q} .

When using a Bloch representation for the electronic eigenstates, the variation of the effective potential, $\delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}}$, connects states of momentum \mathbf{k} with those of momentum $\mathbf{k} + \mathbf{q}$. The Fourier transform of the first order density variation takes the form

$$\delta_{\kappa\alpha}^{\mathbf{q}} n(\mathbf{q} + \mathbf{G}) = -\frac{4}{V} \sum_{\mathbf{k}v} \langle \mathbf{k}v | e^{-i(\mathbf{q}+\mathbf{G})\mathbf{r}} | \Delta_{\kappa\alpha}^{\mathbf{q}}(\mathbf{k}v) \rangle, \quad (43)$$

where V denotes the crystal volume. The quantity appearing on the right hand side is closely related to the first-order variation of the valence state $|\mathbf{k}v\rangle$ and is defined by (see Eq. (37))

$$|\Delta_{\kappa\alpha}^{\mathbf{q}}(\mathbf{k}v)\rangle = \sum_c \frac{|\mathbf{k} + \mathbf{q}c\rangle \langle \mathbf{k} + \mathbf{q}c | \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}} | \mathbf{k}v \rangle}{\epsilon_c(\mathbf{k} + \mathbf{q}) - \epsilon_v(\mathbf{k})}. \quad (44)$$

It is obtained by solving the inhomogeneous linear equations (see Eq. (38))

$$(H_{KS}^{\mathbf{k}+\mathbf{q}} - \epsilon_v(\mathbf{k})) |\Delta_{\kappa\alpha}^{\mathbf{q}}(\mathbf{k}v)\rangle = (P_v^{\mathbf{k}+\mathbf{q}} - 1) \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}} |\mathbf{k}v\rangle. \quad (45)$$

Eqs. (43) and (45) together with (28) constitute a set of equations that is solved self-consistently for a fixed \mathbf{q} to obtain $\delta_{\kappa\alpha}^{\mathbf{q}} n$. As a by-product, $\delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}}$ is also calculated.

The electronic contribution to the dynamical matrix takes the form

$$\delta_{\kappa\alpha}^{\mathbf{q}} \delta_{\kappa'\beta}^{-\mathbf{q}} E = \sum_{\mathbf{G}} [\delta_{\kappa\alpha}^{\mathbf{q}} n(\mathbf{G} + \mathbf{q}) \delta_{\kappa'\beta}^{-\mathbf{q}} v_{\text{ext}}(\mathbf{G} + \mathbf{q}) + \delta_{\kappa\alpha}^{\mathbf{q}} \delta_{\kappa'\beta}^{-\mathbf{q}} v_{\text{ext}}(\mathbf{G})]. \quad (46)$$

Typically, first principles calculations along these lines are performed on a grid of \mathbf{q} -points that form a regular lattice and span the whole Brillouin zone. Discrete Fourier transforms are then applied to interpolate dynamical matrices on arbitrary points in between.

2.3.1 Technical aspects and extensions

The above derivation sketched the main ideas behind the perturbative approach. Practical implementations in existing band structure techniques require a variety of extensions and generalizations, which will be briefly discussed here.

Metals: Originally, the scheme was formulated for non-metallic systems and first applied in the framework of the plane-wave pseudopotential method. An extension to metallic systems has been derived by de Gironcoli, which contains essentially technical modifications related to the appearance of fractional occupation numbers for electronic states with energies close to the Fermi energy [39].

Non-local potentials: The above derivation assumed a local external potential. Modern pseudopotential approaches typically use also non-local forms, for which the above derivation is not strictly valid. It can be modified to include these forms, with the caveat that the dynamical matrix (46) cannot be expressed solely in terms of the density variation anymore but explicitly involves first-order variations of the wavefunctions, too.

Basis-set corrections: In many implementations, the electronic states are expanded in terms of a basis set. If the basis functions do not depend on the ionic positions, as is true of plane waves, the formulation given above is essentially unchanged. However, if the basis set depends explicitly on the position of the ions, it gives rise to additional contributions related to the change of the basis functions under an ionic displacement. These basis set corrections are known as Pulay corrections in the context of force calculations. Similar correction terms occur for methods based on ultrasoft pseudopotentials, as their construction requires the introduction of auxiliary charges centered at ionic sites.

Spin polarization: Extension to spin-polarized DFT is straightforward. The two spin sectors can be treated independently in the perturbation calculation, because the perturbation potential $\delta^q v$ connects states of equal spins only. The dynamical matrix is then given as a sum of contributions from each spin.

Relativistic corrections: Extensions of the semi-relativistic framework, where spin-orbit coupling is neglected, have recently been worked out in the context of pseudopotentials [40, 41]. Here spin-orbit coupling can be easily incorporated by an additive term in the pseudopotential $v_{\text{ext}} \rightarrow v^{SR} + v^{SOC}$. It depends on the ion positions and gives rise to additional terms in δv_{ext} which entangle spatial and spin degrees of freedom.

3 Electron phonon coupling

3.1 Density functional perturbation approach to electron phonon vertex

3.1.1 Form of electron-phonon vertex

In the previous section, we have shown how the dynamics of the ions can be calculated quantitatively within the DFT approach. This was done on the basis of the adiabatic approximation, where the electronic subsystem entered the ionic equation-of-motion via a static screening term

only. Now we go one step further and consider the effect of a dynamical coupling between the electronic and phononic subsystems. Here, only the main ideas to derive the basic form of the electron-phonon vertex are sketched. A more elaborate discussion can be found in the book of Grimvall [42].

Let us consider again the general Hamiltonian of a solid given in Eq. (1) and look at the matrix element

$$\langle n\alpha | \mathcal{H} | n'\alpha' \rangle, \quad (47)$$

where a state $|n\alpha\rangle$ denotes the product of separate electronic and phononic wavefunctions. In the adiabatic approximation, only diagonal elements $n = n'$ and $\alpha = \alpha'$ are present. Non-vanishing off-diagonal elements come from the non-adiabatic terms ΔH in Eq. (6). For the most important one, $\Delta H^{(1)}$, one obtains

$$\langle n\alpha | \Delta H^{(1)} | n'\alpha' \rangle = \int \chi_\alpha^* (\psi_n^* \nabla_{\mathbf{R}} \psi_{n'}) \cdot \nabla_{\mathbf{R}} \chi_{\alpha'} . \quad (48)$$

As before one assumes that all quantities are expanded around the equilibrium positions $\mathbf{R} = \mathbf{R}^0 + \mathbf{u}$ in a fast converging expansion in terms of u . To get an explicit expression, we describe the change in the electronic wavefunction by an effective potential $V(\mathbf{R})$ due to small ionic displacements, giving

$$\int \psi_n^* \nabla_{\mathbf{R}} \psi_{n'} \propto \langle n | \nabla_{\mathbf{R}} V | n' \rangle, \quad (49)$$

where $|n\rangle$ and $|n'\rangle$ denote unperturbed electronic states. The remaining ionic matrix element $\int \chi_\alpha^* \nabla_{\mathbf{R}} \chi_{\alpha'}$ is proportional to the momentum operator, which depends linearly on the phonon creation and annihilation operators.

Thus the off-diagonal matrix elements in (47) describe the probability of emission or absorption of a phonon under a simultaneous excitation in the electronic subsystem. The electronic transition probabilities are determined by the first-order variation of the effective potential $V(\mathbf{R})$ with respect to the ion coordinates as the perturbation operator.

3.1.2 Electron-phonon vertex in density functional perturbation theory

In the context of DFT, the electron-phonon coupling (EPC) matrix elements are defined as transition probabilities of Kohn-Sham states induced by a change in the potential due to a small ion displacement. If one would choose the electron-ion interaction potential, one obtains according to Eq. (42)

$$\langle \mathbf{k} + \mathbf{q}\nu' | \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{ext}} | \mathbf{k}\nu \rangle = -\langle \mathbf{k} + \mathbf{q}\nu' | e^{i\mathbf{q}\mathbf{r}} \sum_l e^{i\mathbf{q}(\mathbf{R}_{l\kappa}^0 - \mathbf{r})} \nabla_{\alpha}^{\mathbf{r}} v_{\kappa}(\mathbf{r} - \mathbf{R}_{l\kappa}^0) | \mathbf{k}\nu \rangle. \quad (50)$$

It is convenient to switch to the normal-mode representation

$$g_{\mathbf{k}+\mathbf{q}\nu', \mathbf{k}\nu}^{(0)\mathbf{q}j} = \sum_{\kappa\alpha} A_{\kappa\alpha}^{\mathbf{q}j} \langle \mathbf{k} + \mathbf{q}\nu' | \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{ext}} | \mathbf{k}\nu \rangle, \quad (51)$$

where $\mathbf{q}j$ denotes the normal modes with momentum \mathbf{q} and mode index j , and the coefficients $A_{\kappa\alpha}^{\mathbf{q}j}$ are defined in (16). This expression describes a rigid shift of the ionic potential, and

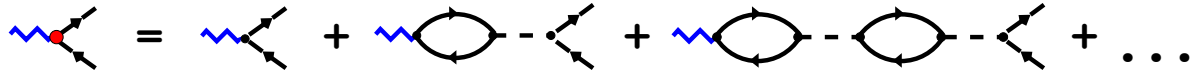


Fig. 1: Diagrammatic representation of the screened electron-phonon vertex within the DFT framework. Blue zigzag lines represent phonons, black lines electron propagators, and the dashed lines the effective electron-electron interaction.

constitutes the so-called *bare* electron-phonon coupling matrix elements. Such a rigid-ion approximation is only justified in cases where all electrons are tightly bound to the ions as in an ionic crystal. For metals, in particular, this approximation typically fails, because it neglects the reaction of the electrons on the disturbance, which tend to screen the perturbation of the potential.

Within DFT the *screened* electron-phonon matrix elements are given by the variation of the effective potential

$$g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}\lambda} = \sum_{\kappa\alpha} A_{\kappa\alpha}^{\mathbf{q}j} \langle \mathbf{k} + \mathbf{q}\nu' | \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}} | \mathbf{k}\nu \rangle. \quad (52)$$

It is instructive to look at it from a many-body perturbation perspective. Fig. 1 shows a diagrammatic representation of the screened vertex. The bare vertex is screened by virtual electron-hole excitations coupled via an effective interaction. From the relationship (34) between the external (bare) and effective (screened) perturbation, we can see that within the DFPT framework, the electron-hole bubble is represented by the charge-susceptibility of the non-interacting Kohn-Sham system (32). The effective interaction is given by the kernel I defined in Eq. (28) and incorporates both the Coulomb interaction and contributions from exchange and correlation.

We have seen in Sec. 2.3 that the first-order variation of the effective potential is calculated as a by-product in the DFPT self-consistent procedure. As the EPC matrix elements contain only this quantity and the unperturbed Kohn-Sham states, they can be calculated by a comparatively small numerical effort after a DFPT cycle is converged. They provide detailed microscopic information about how the coupling depends on the momenta of the electronic and phononic states as well as on the character of the electronic wavefunctions and the displacement pattern of the normal mode, respectively.

The spin dependence is incorporated in the EPC vertex in a straightforward way. In the semi-relativistic framework, when spin-orbit coupling is neglected, spin and spatial degrees-of-freedom are independent, and the perturbation due to the ion displacements does not flip the spin. As a consequence the EPC vertex is spin-diagonal. The EPC vertex in the presence of spin-orbit coupling has a more complex form. In the context of the pseudopotential method, it was shown that a second additive contribution to the perturbation operator appears, $\delta v_{\text{eff}} \rightarrow \delta v_{\text{eff}}^{SR} + \delta v^{SOC}$, which results in off-diagonal spin terms [43].

The EPC matrix elements are the essential ingredients for a numerical approach to a variety of physical properties. In the following I will discuss in some detail (i) renormalization of phonon properties, (ii) phonon-mediated pairing interaction and superconductivity, and (iii) self-energy effects for electronic states. Emphasis will be on how one can connect theoretical predictions with experimental observations to test the accuracy of the first principles EPC matrix elements,

$$\begin{aligned}
(a) \quad \Pi(\omega) &= \text{diagram 1} + \text{diagram 2} - \text{diagram 3} \\
&\quad + \text{diagram 4} - \text{diagram 5} - \text{diagram 6} + \dots \\
(b) \quad \text{Re } \Pi(0) &= \text{diagram 7} \\
(c) \quad \text{Im } \Pi(\omega) &= \text{Im} \text{diagram 8}
\end{aligned}$$

Fig. 2: (a) Diagrammatic representation of the phonon self-energy up to second order in the electron-phonon vertex. Blue zigzag lines represent phonons, black lines electron propagators, and the dashed lines the effective electron-electron interaction. The leading contributions in the limit $\omega \rightarrow 0$ can be summed up with the help of the screened vertex introduced in Fig. 1. They are shown for (b) real part and (c) imaginary part of Π , respectively.

which by themselves are not directly measurable.

In these applications one frequently encounters the problem that one needs EPC matrix elements on momentum grids, which are finer than the ones used in the DFPT cycle. Denser \mathbf{k} grids can be applied in a straightforward way because the calculation of the EPC matrix elements is done separately from the DFPT part, so that only additional Kohn-Sham states need to be calculated. This is a numerically rather cheap procedure. In contrast, the DFPT calculation of $\delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}}$ is much more demanding. Therefore, matrix elements $g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j}$ for intermediate \mathbf{q} are obtained by interpolation techniques. One way is to interpolate $\delta^{\mathbf{q}i} v_{\text{eff}}$ given on a regular \mathbf{q}_i -grid and use exact Kohn-Sham states to evaluate the matrix elements. An alternative route is to use electron and phonon Wannier functions to represent the EPC matrix elements and utilize their spatial localization to interpolate them on very fine momentum grids [44, 45].

3.2 Phonon self-energy and linewidth

On the level of DFPT in the harmonic approximation, phonons are elementary excitations of the lattice, which do not interact and therefore have infinite lifetime. The interaction with electrons results in renormalized quasiparticle properties expressed by the phonon self-energy Π . The renormalized phonon Greens function is obtained from the bare Greens function $D_{0,\mathbf{q}j}(\omega) = 1/(\omega - \omega_{\mathbf{q}j}) - 1/(\omega + \omega_{\mathbf{q}j})$ via the Dyson equation $D^{-1} = D_0^{-1} - \Pi$ as

$$D_{\mathbf{q}j}(\omega) = \frac{2\omega_{\mathbf{q}j}}{\omega^2 - \omega_{\mathbf{q}j}^2 - 2\omega_{\mathbf{q}j}\Pi_{\mathbf{q}j}(\omega)}. \quad (53)$$

For not too large self-energies, the renormalization of the phonons leads to (i) a broadening of the quasiparticle peak of the spectral function, i.e. a finite linewidth, which is proportional to the inverse lifetime, and (ii) to a shift of the peak position. The linewidth is connected to $\text{Im}\Pi$, and the peak shift to $\text{Re}\Pi$.

In the following, we briefly sketch the approach to the phonon renormalization in the DFPT framework. It basically focuses on the lowest-order corrections beyond the adiabatic approximation. The starting point is a diagrammatic representation of the self-energy contributions coming from second-order EPC as sketched in Fig. 2(a). In general, the electron-hole bubble depends on the frequency ω and varies on the scale of electronic energies. These are usually much larger than typical phonon frequencies relevant for the renormalization. Therefore, to lowest order, one can take the static limit $\omega = 0$, and the series of diagrams can be summed up with the help of the screened vertex introduced above in Fig. 2(b). This gives a contribution to $\text{Re}\Pi$ only, which has the form

$$\text{Re}\Pi_{\mathbf{q}j}(0) = \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j} \left(g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{(0)\mathbf{q}j} \right)^* \frac{f(\epsilon_{\mathbf{k}\nu}) - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})}{\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}}. \quad (54)$$

Here $f(\epsilon) = (1 + e^{(\epsilon-\mu)/k_B T})^{-1}$ denotes the Fermi distribution function. Eq. (54) involves both the screened and bare vertices due to fact that the screening enters in a symmetric way, and using the screened vertex on both sides would result in a double counting of the diagrams containing the dashed lines.

It is now instructive to see that this static renormalization is already included in the DFPT procedure. This can be seen by using the definitions of the screened and bare vertices

$$\begin{aligned} \text{Re}\Pi_{\mathbf{q}j}(0) &= \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j} \left(g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{(0)\mathbf{q}j} \right)^* \frac{f(\epsilon_{\mathbf{k}\nu}) - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})}{\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}} \\ &= \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} \sum_{\kappa\alpha\kappa'\beta} A_{\kappa\alpha}^{\mathbf{q}j} A_{\kappa'\beta}^{-\mathbf{q}j} \langle \mathbf{k} + \mathbf{q}\nu' | \delta_{\kappa\alpha}^{\mathbf{q}} v_{\text{eff}} | \mathbf{k}\nu \rangle \langle \mathbf{k}\nu | \delta_{\kappa'\beta}^{-\mathbf{q}} v_{\text{ext}} | \mathbf{k} + \mathbf{q}\nu' \rangle \\ &\quad \times \frac{f(\epsilon_{\mathbf{k}\nu}) - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})}{\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}} \\ &= \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} \sum_{\kappa\alpha\kappa'\beta} A_{\kappa\alpha}^{\mathbf{q}j} A_{\kappa'\beta}^{-\mathbf{q}j} \int d^3r \delta_{\kappa\alpha}^{\mathbf{q}} n(\mathbf{r}) \delta_{\kappa'\beta}^{-\mathbf{q}} v_{\text{ext}}(\mathbf{r}). \end{aligned} \quad (55)$$

In the last step the linear-response expression (30) for the first-order variation of the electronic density was applied. Comparison with Eq. (46) shows that this term corresponds to the first contribution to the dynamical matrix which comes from the variation of the density. Thus this renormalization is already taken into account on the level of DFPT.

The situation is different for $\text{Im}\Pi$. This is a true non-adiabatic property, and one has to go beyond the static approximation. The dominant contribution in the limit $\omega \rightarrow 0$ is obtained by replacing in each term of the series in Fig. 2(a) one electron-hole bubble by its imaginary part and taking the static limit for all others. Then again, the series can be summed up, but now both vertices in the diagram are screened (Fig. 2(c)). This leads to the following expression for the linewidth (half-width at half maximum)

$$\begin{aligned} \gamma_{\mathbf{q}j} &= -2 \text{Im}\Pi_{\mathbf{q}j}(\omega_{\mathbf{q}j}) \\ &= 2\pi \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} |g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j}|^2 (f(\epsilon_{\mathbf{k}\nu}) - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})) \delta(\omega_{\mathbf{q}j} + (\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'})). \end{aligned} \quad (56)$$

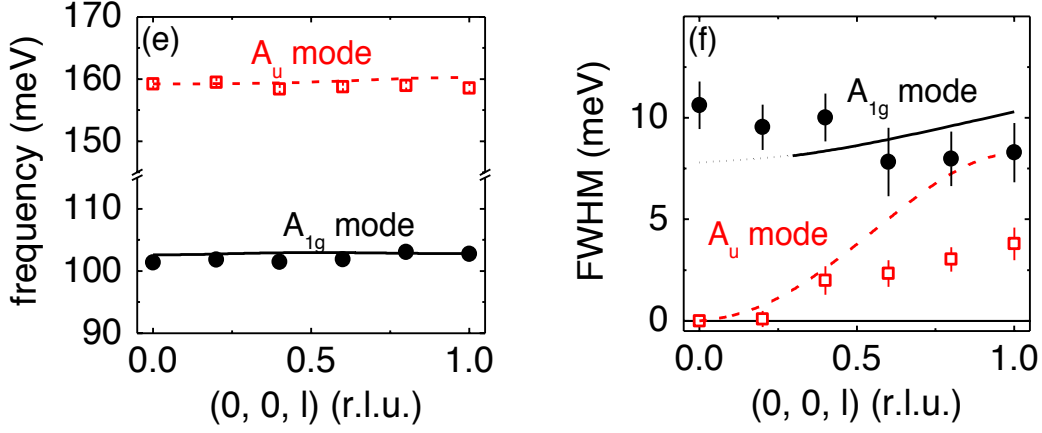


Fig. 3: Phonon dispersion (left panel) and linewidths (right panel) for two high-frequency branches of $\text{YNi}_2\text{B}_2\text{C}$ along the $[001]$ direction. Shown are inelastic neutron scattering results (symbols) together with first principles predictions (lines) [47].

This expression contains the T -dependence via the Fermi distribution functions f . It can be further simplified as long as the electronic structure has no peculiarities on the scale of phonon energies. The δ -function forces the energy difference $\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}$ to be small, hence the difference of the Fermi distribution functions can be approximated with the help of its energy derivative $f' = df/d\epsilon$

$$f(\epsilon_{\mathbf{k}\nu}) - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'}) \approx f'(\epsilon_{\mathbf{k}\nu})(\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}) \rightarrow -f'(\epsilon_{\mathbf{k}\nu})\omega_{\mathbf{q}j}. \quad (57)$$

Finally, the phonon frequency is neglected in the δ function.

At low temperatures, the energy derivative of the Fermi distribution function is strongly peaked at the Fermi energy. In the limit $T \rightarrow 0$, it can be replaced by $f'(\epsilon_{\mathbf{k}\nu}) \rightarrow -\delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F)$. We finally arrive at an expression for the EPC-induced phonon linewidth valid in the limit $T \rightarrow 0$

$$\gamma_{\mathbf{q}j} = 2\pi\omega_{\mathbf{q}j} \frac{1}{N_k} \sum_{\mathbf{k}\nu\nu'} |g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F) \delta(\epsilon_{\mathbf{k}+\mathbf{q}\nu'} - \epsilon_F). \quad (58)$$

It contains only quantities that are available within the DFPT approach to the EPC. The derivation of (58) was first given by Allen [46]. This form of the linewidth is most often used in first-principles calculations. One must, however, be aware of the approximations underlying its derivation. It is only valid in the limit $T \rightarrow 0$ and also breaks down in the limit $\mathbf{q} \rightarrow 0$ for metals because intraband scattering events involve arbitrarily small energy differences $\epsilon_{\mathbf{k}\nu} - \epsilon_{\mathbf{k}+\mathbf{q},\nu}$, and the phonon frequency cannot be neglected anymore.

Phonon linewidths can be measured by, e.g., inelastic neutron or x-ray scattering experiments. However, when comparing theory and experiment, one should keep in mind that the above formula only represents the contribution to the linewidth coming from EPC. Experimentally, one needs to separate it from other possible contributions to the linewidth, most importantly those related to anharmonic decay processes, which often is not easy to achieve.

An example is shown for two high-frequency modes of the superconductor $\text{YNi}_2\text{B}_2\text{C}$ in Fig. 3. Measurements were done at low T , where anharmonic contributions are supposed to play a

minor role. The calculations do predict the size and momentum dependence of the linewidths reasonably well, demonstrating that even for compounds with rather complex lattice structure these calculations are reliable.

As we will see in the discussion of superconductivity given below, the same expression (58) enters the pairing properties. As such Eq. (58) provides a link between the pairing strength of a phonon mode and the linewidth, the latter being an experimentally accessible quantity.

As discussed in Sec. 3.1, the concept of electron-phonon coupling goes beyond the adiabatic approximation, and we leave the firm base of DFT. The above derivation showed, however, that to a first approximation the phonon linewidth can still be calculated within the DFPT scheme. To include further non-adiabatic corrections as for, e.g., the frequency renormalization, a more general framework like the time-dependent DFT is required. Such a generalization was outlined recently by Saitta *et al.* [48].

3.3 Phonon mediated pairing interaction and superconductivity

Superconductivity is a macroscopic quantum phenomenon of the electron system. Its origin lies in an instability of the Fermi liquid state and leads to a new ground state of correlated paired electrons (Cooper pairs). In their seminal paper, Bardeen, Cooper, and Schrieffer (BCS) [49] have shown that this state is stabilized when there is a small attractive interaction among two electrons. Such an attractive interaction is always provided by the electron-phonon coupling, which thus represents a natural source for pairing in any metal. EPC is known to be the pairing mechanism in most superconductors, which are commonly termed classical superconductors to distinguish them from more exotic materials where other types of pairing mechanism are suspected.

The BCS theory treated the EPC only in a simplified form appropriate for the weak coupling limit. Soon after a more complete theory was worked out, applying many-body techniques (for a review see, e.g., Scalapino [50]). The resulting Eliashberg theory [51] extends the framework of BCS into the strong coupling regime and allows a quantitative prediction of many properties of the superconducting state. An important property of the superconducting state is that the quasiparticle spectrum is gapped. The size of the gap plays the role of an order parameter. It is determined from a self-consistent solution of a set of equations which generalize the BCS gap equations. An important feature of these so-called Eliashberg gap equations is that only normal-state properties enter, which specify a particular material. These include details about the electronic structure and the phonon-mediated pairing interaction, quantities which are readily accessible within the first principles approach to EPC outlined above.

As the Eliashberg theory will be presented in detail in another lecture of this Autumn School, I will focus here on the procedure to calculate the effective electron-electron interaction. The physical process behind the phonon-mediated interaction is the exchange of a phonon between two electrons, shown schematically in Fig. 4. This translates into the so-called Eliashberg function

$$\alpha^2 F_{\mathbf{k}\nu, \mathbf{k}'\nu'}(\omega) = N(\epsilon_F) \frac{1}{N_q} \sum_{\mathbf{q}j} |g_{\mathbf{k}'\nu', \mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\omega - \omega_{\mathbf{q}j}). \quad (59)$$

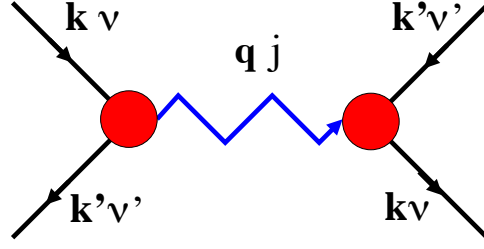


Fig. 4: Diagrammatic representation of the effective electron-electron interaction mediated by the exchange of a phonon.

Here $N(\epsilon_F)$ is the electronic density-of-states at the Fermi energy per spin. The sum extends over all phonon modes mediating the interaction, which also determine its frequency dependence. Both coupling vertices are represented by screened matrix elements, and its implicit momentum conservation restricts the sum to $\mathbf{q} = \mathbf{k}' - \mathbf{k}$. For the superconducting pairing, this interaction is most effective for electronic states with energies $|\epsilon_{\mathbf{k}\nu} - \epsilon_F| \leq \omega_{\text{phonon}}$. Thus, in practice one evaluates (59) only for states right at the Fermi surface (for an elaborate discussion of the underlying assumptions and approximations see [52]).

Most superconductors exhibit surprisingly isotropic superconducting gaps. The reason is that defects, which are always present in real materials, tend to wash out the momentum dependence of the interaction. In many cases, it is sufficient to consider the Fermi surface average leading to the isotropic Eliashberg function $\alpha^2 F(\omega) = \sum_{\mathbf{k}\nu, \mathbf{k}'\nu'} w_{\mathbf{k}\nu} w_{\mathbf{k}'\nu'} \alpha^2 F_{\mathbf{k}\nu, \mathbf{k}'\nu'}(\omega)$ with $w_{\mathbf{k}\nu} = \delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F) / N(\epsilon_F)$. Taking momentum conservation into account, this is typically cast into the form

$$\alpha^2 F(\omega) = \frac{1}{N(\epsilon_F)} \frac{1}{N_q} \sum_{\mathbf{q}j, \mathbf{k}\nu\nu'} |g_{\mathbf{k}+\mathbf{q}\nu', \mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\omega - \omega_{\mathbf{q}j}) \delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F) \delta(\epsilon_{\mathbf{k}+\mathbf{q}\nu'} - \epsilon_F). \quad (60)$$

Within Eliashberg theory, important characteristics like the superconducting transition temperature often depend on integrated quantities only. One such quantity is the isotropic coupling constant defined by

$$\lambda = 2 \int d\omega \frac{\alpha^2 F(\omega)}{\omega}, \quad (61)$$

which is the dimensionless measure of the average strength of the coupling. Commonly, values larger than 1 are characterized as strong coupling. The factor $1/\omega$ in the integral indicates that low-energy modes are generally more effective than high-energy modes in mediating the pairing.

At this stage it is useful to make a connection to the expression for the phonon linewidth derived in the limit $T \rightarrow 0$. Using Eq. (58), the isotropic Eliashberg function can be written as

$$\alpha^2 F(\omega) = \frac{1}{2\pi N(\epsilon_F)} \frac{1}{N_q} \sum_{\mathbf{q}j} \frac{\gamma_{\mathbf{q}j}}{\omega_{\mathbf{q}j}} \delta(\omega - \omega_{\mathbf{q}j}) \quad (62)$$

and the isotropic coupling constant as

$$\lambda = \frac{1}{\pi N(\epsilon_F)} \frac{1}{N_q} \sum_{\mathbf{q}j} \frac{\gamma_{\mathbf{q}j}}{\omega_{\mathbf{q}j}^2}. \quad (63)$$

The dimensionless prefactor $\gamma_{\mathbf{q}j}/\omega_{\mathbf{q}j}$ in (62) can be interpreted as a measure of the coupling due to an individual phonon mode. The Eliashberg function is then given as a sum over all phonon branches and averaged over phonon momentum.

Commonly, DFPT based calculations of the Eliashberg function are done by first calculating $\gamma_{\mathbf{q}j}$ and then performing the sum over the phonon spectrum. The appearance of a product of δ -functions in the expression (58), however, requires a careful numerical approach. Usually the δ -functions are replaced by smoother functions like Gaussians, but to reach convergence the \mathbf{k} -summation has to be carried out on meshes which are denser than the one used for the calculation of the phonon properties.

The strong-coupling superconductor Pb illustrates this general approach well because the effects of EPC are especially pronounced and relativistic corrections are necessary for a satisfactory quantitative description. Fig. 5(a) shows the phonon dispersion along high-symmetry directions, comparing semi-relativistic (SR) results and those including the spin-orbit coupling (SOC) with data from inelastic neutron-scattering experiments. The experimental scale of phonon frequencies, and in particular the pronounced anomalous dips observed at various points along the dispersion curves are fingerprints of strong EPC. They are much better reproduced when SOC is included, which in addition improves the whole spectrum by lowering the frequencies as compared to the SR calculation. Phonon linewidths shown in Fig. 5(b) for the same branches exhibit a strong variation as a function of momentum. The values are generally larger for the SOC calculation, indicating that the matrix elements are enhanced significantly. This is reflected in an enhanced $\alpha^2 F$ shown in Fig. 5(c), which much better agrees with the Eliashberg function derived from tunneling spectroscopy experiments. The isotropic coupling constant λ is increased from 1.08 to 1.56 by SOC, which matches well the experimental value of 1.55. A closer look at the different contributions to λ reveals that about half of this increase originates in the softening of the spectrum (due to the factor $1/\omega$ in the expression for λ), while the other half comes from an increase of the coupling matrix elements. This example demonstrates that SOC can substantially modify the EPC [43].

Anisotropic superconducting states can be handled using the full momentum dependence of the Eliashberg function (59). This has been done rarely in the past, as the fully anisotropic gap equations are difficult to solve. A special class of anisotropic superconductors are multiband superconductors, which possess several Fermi surface sheets. The superconducting gap can vary among the different sheets but is approximately isotropic on a single sheet. In this case, a partially averaged pairing function is appropriate

$$\alpha^2 F_{\nu\nu'}(\omega) = \frac{1}{N(\epsilon_F)} \frac{1}{N_q} \sum_{\mathbf{q}j,\mathbf{k}} |g_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\omega - \omega_{\mathbf{q}j}) \delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F) \delta(\epsilon_{\mathbf{k}+\mathbf{q}\nu'} - \epsilon_F). \quad (64)$$

The isotropic Eliashberg function is replaced by a matrix describing intraband and interband pairing contributions.

A textbook example of such a multiband superconductor is MgB_2 . Here two types of electronic states are present at the Fermi level: σ and π states, which are derived mainly from the boron p states. Calculations of the band-resolved Eliashberg functions shown in Fig. 6 revealed that the

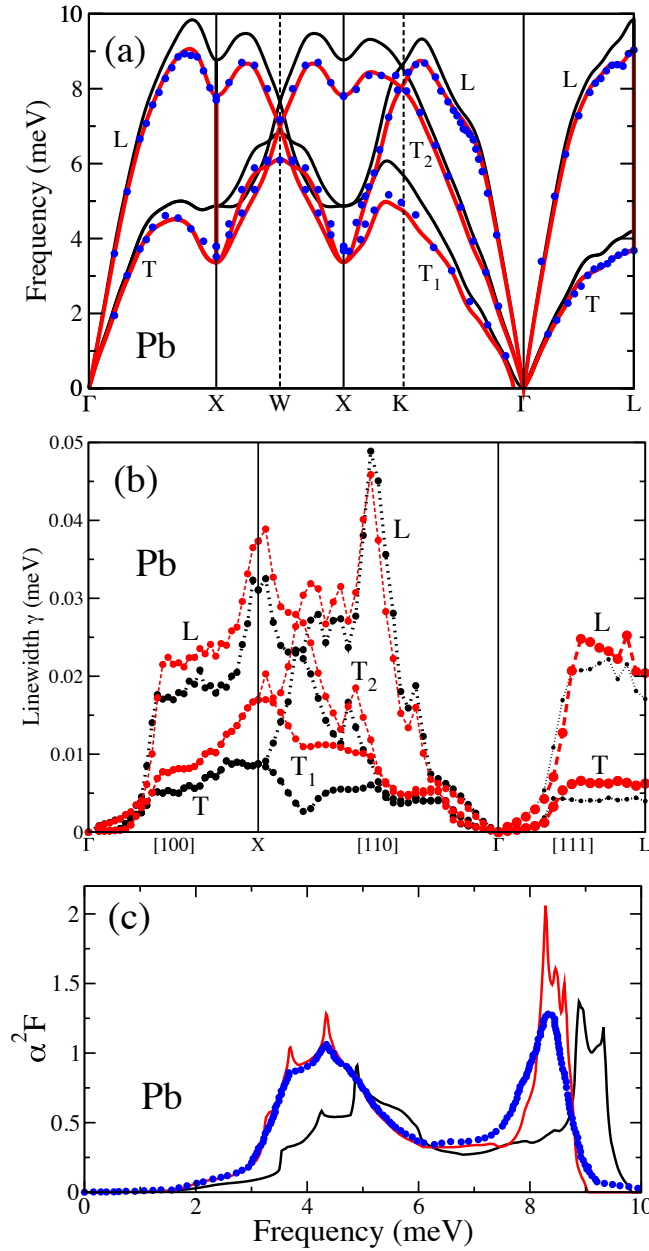


Fig. 5: Phonon and electron-phonon coupling properties for the elemental superconductor Pb. Figures show (a) the phonon dispersion, (b) the mode-dependent linewidth, and (c) the Eliashberg function (c). Black lines indicate results of calculations without spin-orbit coupling, and red lines with spin-orbit coupling included. For comparison, blue symbols show experimental results for (a) phonon frequencies obtained from inelastic neutron scattering experiments [53] and (c) the Eliashberg function extracted from tunneling spectroscopy data [54].

pairing interaction is predominantly driven by the intraband σ - σ contribution. It originates from a strong coupling to in-plane B vibrations which drive a large softening of branches connected to the E_{2g} modes (left panel in Fig. 6). This peculiar pairing interaction leads to a superconducting state with gaps of different magnitude for the σ and π Fermi surfaces whose signature could be found, e.g., in specific heat measurements. Tunneling spectroscopy gave strong support for the predicted multiband pairing spectrum [55], too.

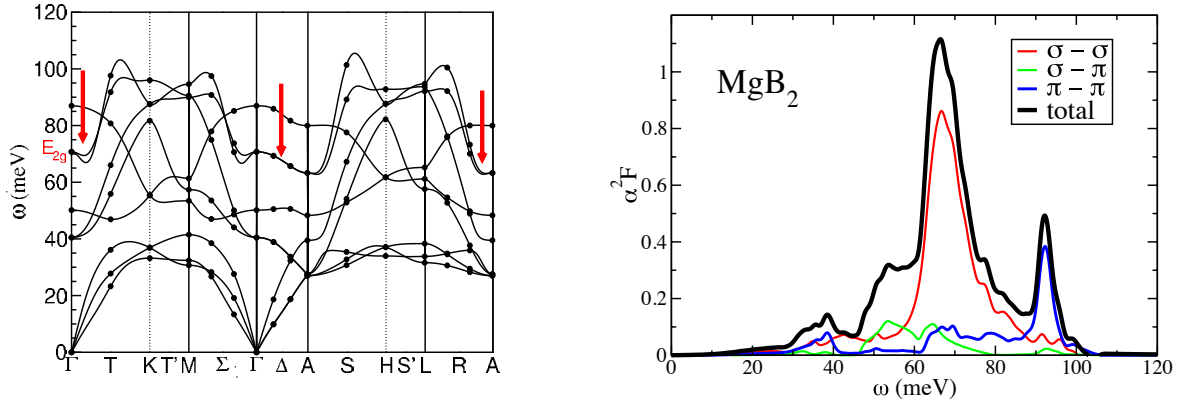


Fig. 6: Left panel: Calculated phonon dispersion of the multiband superconductor MgB_2 . The large frequency softening of the E_{2g} -related branches due to a very strong EPC is indicated by red arrows. Right panel: Calculated total and band-resolved Eliashberg functions of MgB_2 demonstrating that the dominant coupling originates from a large intraband σ - σ pairing interaction.

3.4 Electron self-energy effects

Another important physical consequence of EPC is the renormalization of electronic quasiparticles. This has particularly profound consequences for metals, as it strongly modifies electronic states whose energy distance to the Fermi level is of the order of the phonon energies. Although this is a small energy compared to typical electronic scales, it significantly influences Fermi surface related properties like transport or thermodynamics.

Nowadays, there are a variety of experimental techniques that can probe the properties of electronic quasiparticles in fine detail. A well known example is angle-resolved photoemission spectroscopy (ARPES), which essentially measures the quasiparticle spectral function of occupied states, while unoccupied states can be accessed by, e.g., pump-probe experiments. Such studies provide information about the energy and momentum dependence of the renormalization due to EPC. As these techniques are rather surface sensitive, most applications were devoted to surface electronic states (for a recent review, see [56]).

The renormalization is described by the electron self-energy, which via the Dyson equation enters the renormalized electronic Green's function as

$$G(\mathbf{k}\nu, \epsilon) = \left(\epsilon - (\epsilon_{\mathbf{k}\nu} - \mu) - \Sigma(\mathbf{k}\nu, \epsilon) \right)^{-1}. \quad (65)$$

For not-too-large self-energies, the spectral function $A_{\mathbf{k}\nu}(\epsilon) = -\text{Im}G(\mathbf{k}\nu, \epsilon + i\delta)$ possesses a well defined peak at a shifted quasiparticle energy determined by the real part of Σ

$$\bar{\epsilon}_{\mathbf{k}\nu} = \epsilon_{\mathbf{k}\nu} - \text{Re}\Sigma(\mathbf{k}\nu, \bar{\epsilon}_{\mathbf{k}\nu}). \quad (66)$$

The quasiparticle acquires a finite lifetime leading to a linewidth (FWHM)

$$\Gamma_{\mathbf{k}\nu} = -2\text{Im}\Sigma(\mathbf{k}\nu, \bar{\epsilon}_{\mathbf{k}\nu}) \quad (67)$$

that is determined by the imaginary part.

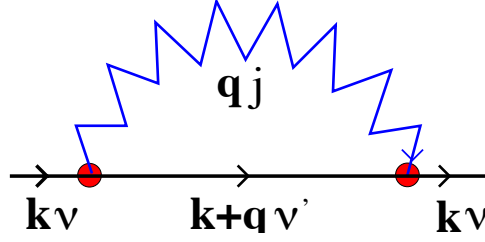


Fig. 7: Diagrammatic representation of the lowest-order contribution to the electron self-energy from the electron-phonon coupling.

The relevant lowest order diagram due to EPC is shown in Fig. 7. It describes a virtual exchange of a phonon. Note that in this case both coupling vertices are screened. This leads to the following expression

$$\Sigma(\mathbf{k}\nu, \epsilon) = \frac{1}{N_q} \sum_{\mathbf{q}j} \sum_{\nu'} |g_{\mathbf{k}+\mathbf{q}\nu', \mathbf{k}\nu}^{\mathbf{q}j}|^2 \left(\frac{b(\omega_{\mathbf{q}j}) + f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})}{\epsilon + \omega_{\mathbf{q}j} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'} + i\delta} + \frac{b(\omega_{\mathbf{q}j}) + 1 - f(\epsilon_{\mathbf{k}+\mathbf{q}\nu'})}{\epsilon - \omega_{\mathbf{q}j} - \epsilon_{\mathbf{k}+\mathbf{q}\nu'} + i\delta} \right). \quad (68)$$

The T dependence now enters via both Fermi and Bose distribution functions, $f(\epsilon)$ and $b(\omega) = [e^{\omega/k_B T} - 1]^{-1}$, respectively.

From Eq. (68) one can readily derive the expression for the imaginary part. It is convenient to introduce two spectral functions which depend explicitly on the electronic state ($\mathbf{k}\nu$)

$$\alpha^2 F_{\mathbf{k}\nu}^{\pm}(\epsilon, \omega) = \frac{1}{N_q} \sum_{\mathbf{q}j} \delta(\omega - \omega_{\mathbf{q}j}) \sum_{\nu'} |g_{\mathbf{k}+\mathbf{q}\nu', \mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\epsilon - \epsilon_{\mathbf{k}+\mathbf{q}\nu'} \pm \omega). \quad (69)$$

Then the imaginary part can be cast in the form

$$\text{Im}\Sigma_{\mathbf{k}\nu}(\epsilon) = -\pi \int_0^\infty d\omega \left\{ \alpha^2 F_{\mathbf{k}\nu}^+(\epsilon, \omega) [b(\omega) + f(\omega + \epsilon)] + \alpha^2 F_{\mathbf{k}\nu}^-(\epsilon, \omega) [b(\omega) + f(\omega - \epsilon)] \right\}. \quad (70)$$

The appearance of δ -functions in (69) allows a straightforward numerical evaluation of Eq. (70). In contrast, a direct calculation of $\text{Re}\Sigma(\mathbf{k}\nu, \epsilon)$ via Eq. (68) requires a summation over many intermediate electronic states, which converges slowly. Therefore, in practice, one obtains $\text{Re}\Sigma$ by making use of the Kramers-Kronig relation

$$\text{Re}\Sigma(\mathbf{k}\nu, \epsilon) = \frac{1}{\pi} \int d\epsilon' \frac{\text{Im}\Sigma(\mathbf{k}\nu, \epsilon')}{\epsilon - \epsilon'}. \quad (71)$$

From (70) one readily obtains an expression for the EPC-induced quasiparticle linewidth

$$\Gamma_{\mathbf{k}\nu} = 2\pi \int_0^\infty d\omega \left\{ \alpha^2 F_{\mathbf{k}\nu}^+(\bar{\epsilon}_{\mathbf{k}\nu}, \omega) [b(\omega) + f(\omega + \bar{\epsilon}_{\mathbf{k}\nu})] + \alpha^2 F_{\mathbf{k}\nu}^-(\bar{\epsilon}_{\mathbf{k}\nu}, \omega) [b(\omega) + f(\omega - \bar{\epsilon}_{\mathbf{k}\nu})] \right\}. \quad (72)$$

The two terms in this expression represent two different scattering processes, which are sketched in Fig. 8. When a quasiparticle hole is created at the state ($\mathbf{k}\nu$), electrons can scatter from states with higher or lower energies, respectively. To conserve the total energy, the first process is connected with the emission of a phonon and is described by $\alpha^2 F_{\mathbf{k}\nu}^-$. On the other hand, $\alpha^2 F_{\mathbf{k}\nu}^+$ describes the scattering of an electron from a lower-energy state with the simultaneous

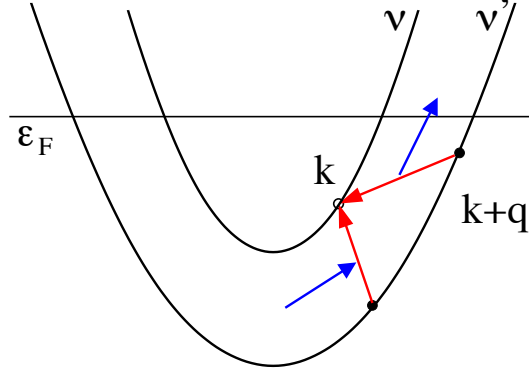


Fig. 8: Illustration of the scattering processes contributing to the self-energy of a hole quasiparticle with momentum \mathbf{k} and band index ν . Electrons (red lines) can scatter virtually from states with higher or lower energies under simultaneous emission or absorption of a phonon (blue lines), respectively.

absorption of a phonon. A similar interpretation can be given if a quasiparticle is created at energies $\epsilon > \epsilon_F$.

While the emission and absorption spectra differ in principle, it is often found in practice that the differences are small. This is due to the fact that the electronic energy scale is typically much larger than the phonon energies. Therefore, emission or absorption of a phonon mainly changes the momentum of the electron, while the energy change is negligible. In this case the phonon energy in the δ -function of (69) can be dropped resulting in $\alpha^2 F_{\mathbf{k}\nu}^\pm \approx \alpha^2 F_{\mathbf{k}\nu}$, with the spectral function

$$\alpha^2 F_{\mathbf{k}\nu}(\epsilon, \omega) = \frac{1}{N_q} \sum_{\mathbf{q}j} \delta(\omega - \omega_{\mathbf{q}j}) \sum_{\nu'} |g_{\mathbf{k}+\mathbf{q}\nu', \mathbf{k}\nu}^{\mathbf{q}j}|^2 \delta(\epsilon - \epsilon_{\mathbf{k}+\mathbf{q}\nu'}). \quad (73)$$

This so-called quasielastic approximation leads to the simplified expression of the linewidth

$$\Gamma_{\mathbf{k}\nu} = \pi \int_0^\infty d\omega \left\{ \alpha^2 F_{\mathbf{k}\nu}(\bar{\epsilon}_{\mathbf{k}\nu}, \omega) [2b(\omega) + f(\omega + \bar{\epsilon}_{\mathbf{k}\nu}) + f(\omega - \bar{\epsilon}_{\mathbf{k}\nu})] \right\}. \quad (74)$$

The coupling parameter is defined via

$$\lambda_{\mathbf{k}\nu} = 2 \int d\omega \frac{\alpha^2 F_{\mathbf{k}\nu}(\epsilon_{\mathbf{k}\nu}, \omega)}{\omega}. \quad (75)$$

This dimensionless quantity characterizes the interaction strength of a specific electronic state to the whole phonon spectrum, and depends both on the momentum and band character of the electronic state. Numerical evaluation of the sum in (73) requires typically rather dense \mathbf{q} -meshes to reach convergence. To get the related matrix elements one resorts to interpolation techniques mentioned in Sec. 3.1.2.

The Eliashberg function discussed in Sec. 3.3 is related to this state-dependent spectral function via appropriate momentum averages at the Fermi energy. For the isotropic Eliashberg function the relation $\alpha^2 F(\omega) = \sum_{\mathbf{k}\nu} w_{\mathbf{k}\nu} \alpha^2 F_{\mathbf{k}\nu}(\epsilon_F, \omega)$ holds with weights $w_{\mathbf{k}\nu} = \delta(\epsilon_{\mathbf{k}\nu} - \epsilon_F)/N(\epsilon_F)$, and the isotropic coupling constant is given by $\lambda = \sum_{\mathbf{k}\nu} w_{\mathbf{k}\nu} \lambda_{\mathbf{k}\nu}$.

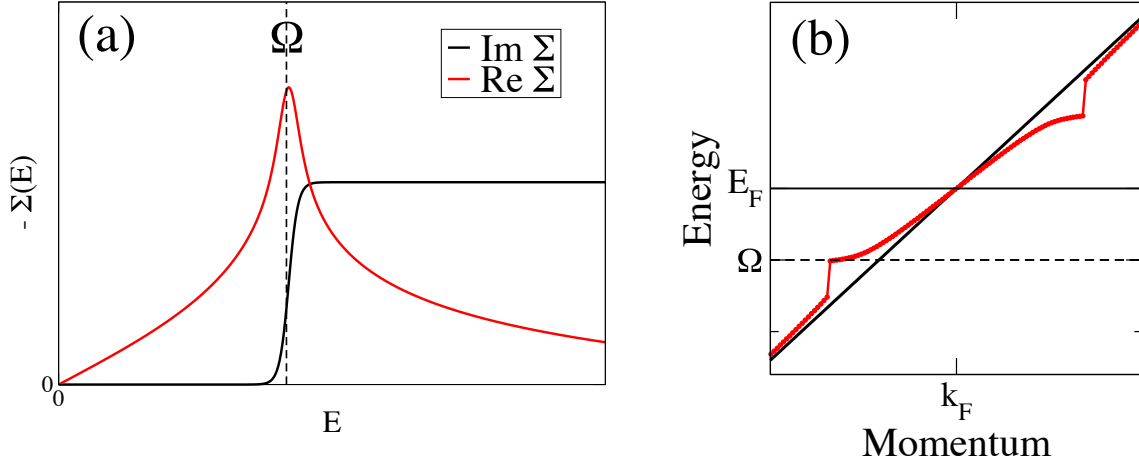


Fig. 9: Illustration of the renormalization of an electronic band for a model coupling to an Einstein-type phonon branch with energy Ω . (a) Real and imaginary part of the electron self-energy. (b) Renormalized quasiparticle dispersion, showing a kink at the phonon frequency.

Experimentally, the coupling strength of individual electronic states can be probed in two different ways. The first is related to the T -dependence of the linewidth. In Eq. (74), it is contained solely in the Bose and Fermi distribution functions. One can easily derive its behavior for the two limiting cases, $T \rightarrow 0$ and large T . For $T \rightarrow 0$, $n(\omega)$ vanishes and one obtains

$$\Gamma_{\mathbf{k}\nu} \rightarrow 2\pi \int_0^{\omega_{\max}} d\omega \alpha^2 F_{\mathbf{k}\nu}(\bar{\epsilon}_{\mathbf{k}\nu}, \omega). \quad (76)$$

With increasing temperature, the linewidth increases for all energies. For temperatures larger than the maximum phonon frequency, this T -dependence becomes linear, and its slope is determined by the average coupling parameter defined above

$$\Gamma_{\mathbf{k}\nu} \approx 2\pi \lambda_{\mathbf{k}\nu} k_B T. \quad (77)$$

This relationship has been widely used to extract $\lambda_{\mathbf{k}\nu}$ from measurements of $\Gamma_{\mathbf{k}\nu}(T)$ for surface electronic states. The analysis is, however, complicated by the fact that the measured linewidth also contains contributions from other decay channels, most noticeably due to electron-electron interaction and due to elastic scattering at defects.

A second route to connecting theory and experiment are ARPES measurements of the spectral function of electronic bands close to the Fermi energy. They provide more direct information about the renormalization and the underlying self-energy. It is instructive to first look at a simple model where only a single dispersionless phonon branch is involved in the coupling, i.e. $\alpha^2 F_{\mathbf{k}\nu}(\omega)$ possesses one δ -type peak at the phonon frequency Ω (Einstein model). Fig. 9(a) shows the resulting self-energy for $T \rightarrow 0$. Its imaginary part exhibits a step at Ω because, for $\omega < \Omega$, no phonons are available to promote the scattering, while for $\omega > \Omega$ this scattering channel is opened. The real part of the self-energy exhibits a peak at the phonon frequency, which shows that the largest shifts of the quasiparticle peaks occur for electronic states with $\bar{\epsilon} - \epsilon_F \approx \Omega$. As a consequence, the renormalized dispersion shown in Fig. 9(b) exhibits a kink-like structure right at the phonon energy.

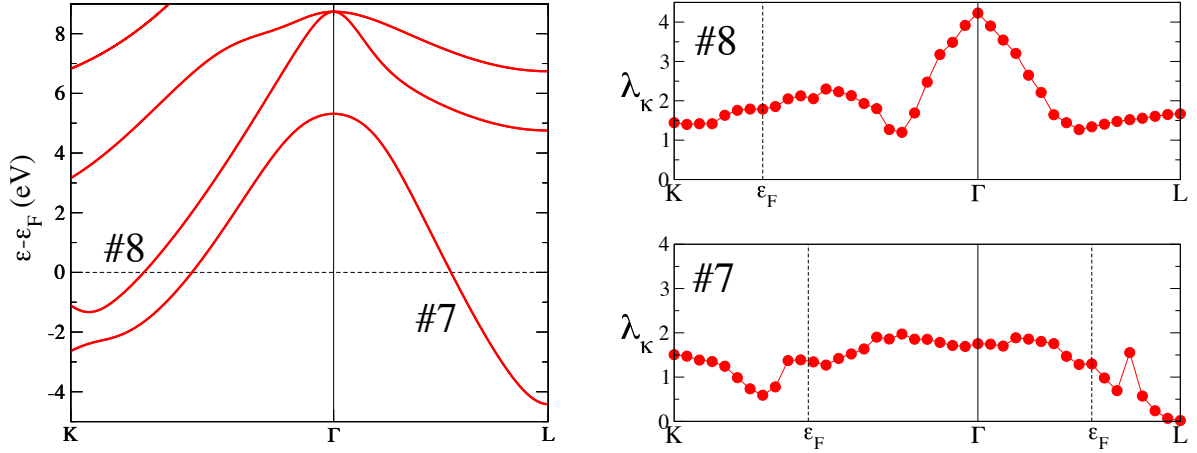


Fig. 10: *Electronic-state dependent coupling constants for Pb. Left panel: band structure along the high-symmetry directions $K\Gamma$ and ΓL . Right panels: Momentum-dependent coupling constants for the two bands crossing the Fermi level. Spin-orbit coupling was taken into account in the calculations.*

In the general case, many phonon modes of different energies contribute to $\alpha^2 F_{\mathbf{k}\nu}(\omega)$ resulting in a broadened self-energy and kink structure. The coupling parameter is related to the slope of $\text{Re}\Sigma$ at the Fermi energy, and the exact relation

$$\lambda_{\mathbf{k}\nu} = \left. \frac{\partial \text{Re}\Sigma(\mathbf{k}\nu, \epsilon)}{\partial \epsilon} \right|_{\epsilon=\epsilon_F, T=0}, \quad (78)$$

holds. It has been used to extract the coupling parameter from measurements of the renormalized electronic bands.

Application of the first principles approach to a real system is exemplified for the case of lead [57]. Fig. 10 shows the results for the coupling constant of electronic states for two bands crossing the Fermi surface. One can observe that $\lambda_{\mathbf{k}\nu}$ can be significantly smaller or larger than the Fermi-surface average of $\lambda = 1.56$ discussed in Sec. 3.3. The variations partly reflect that the availability of intermediate states contributing to the renormalization depends on the binding energy and partly result from a momentum dependence of the EPC matrix elements.

Connection to experiment can be established by comparing the self-energy of states at the Fermi level. Fig. 11 shows data from angle-resolved photoemission spectroscopy experiments for the state of band #8 at the crossing with the Fermi level along the $K\text{-}\Gamma$ direction. The calculations describe the data rather satisfactorily. For this state they predict a coupling constant of $\lambda_k = 1.79$, which is 15% larger than the Fermi-surface average.

The above-given analysis rests on a simplified solution of the Dyson equation which is justified as long as the coupling does not become too large. A generalized treatment has been presented by Eiguren *et al.* [59] that employs a self-consistent solution of the Dyson equation in the complex plane. They showed that for larger coupling the renormalization process becomes much more involved. It can lead to a complex structure of the electronic spectral function, indicating a break-down of the quasiparticle picture.

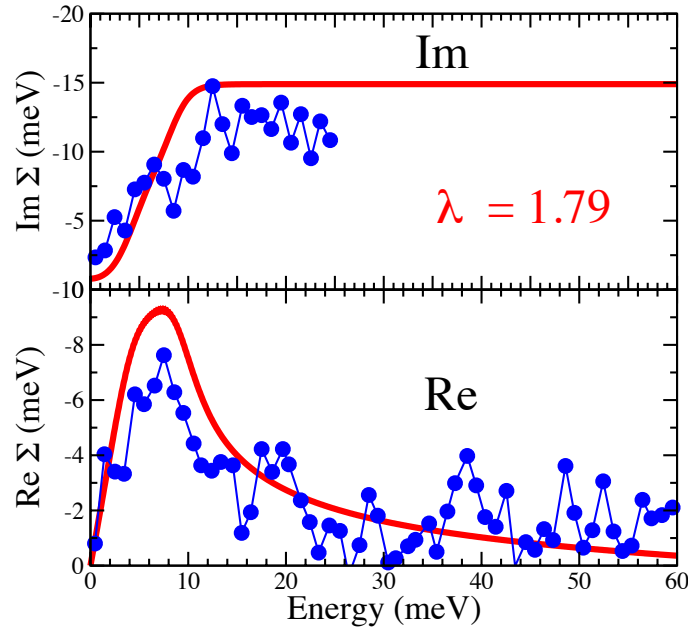


Fig. 11: *Electronic self-energy for Pb. Compared are calculated and measured results for the electronic state at the Fermi-surface crossing of band #8 along the $K\Gamma$ direction (see Fig. 10). Calculations (red lines) include spin-orbit coupling. Angle-resolved photoemission data (blue symbols) were taken from Reinert et al. [58].*

4 Summary

In this tutorial, I have given an introduction to the first principles approach to phonons and electron-phonon interaction within the linear-response framework of density functional theory, the so-called density functional perturbation theory. We have seen that adiabatic lattice vibrations are linked directly to electronic ground-state properties, and their properties can be calculated efficiently within the DFPT scheme. In contrast, electron-phonon coupling is a truly non-adiabatic property. Nevertheless, DFPT provides a simple way to quantify the screened electron-phonon vertex, which is one of the central quantities determining physical observables like phonon linewidth, electron renormalization, or phonon-mediated pairing interaction. This approach is valid as long as the electronic structure does not possess peculiarities on the energy scale of phonons.

This technique has been applied in recent years to a large variety of compounds. Its most common use was devoted to quantitative estimates of the pairing interaction for superconductors, providing valuable information about the basic question of the pairing mechanism. Measurements of quasiparticle renormalization are a more direct tool to gain information about the energy and momentum dependence of the electron-phonon vertex, and offer the possibility to check the accuracy of the DFPT predictions in greater detail. This route should be pursued more systematically in the future. A current challenge in this field is to incorporate more sophisticated treatments of electron correlations in order to extend applications to systems where strong correlations play a crucial role.

References

- [1] M. Born and J.R. Oppenheimer, *Ann. Physik* **84**, 457 (1927)
- [2] G.V. Chester and A. Houghton, *Proc. Phys. Soc.* **73**, 609 (1959)
- [3] M. Born and K. Huang: *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1954)
- [4] A.A. Maradudin, E.W. Montroll, G.H. Weiss, and I.P. Ipatova, in *Solid State Physics, Supplement 3* (Academic Press, New York, 1971), p. 1
- [5] H. Boettger: *Principles of the Theory of Lattice Dynamics* (Physik Verlag, Weinheim, 1983)
- [6] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B 864 (1964)
- [7] W. Kohn and L.J. Sham, *Phys. Rev.* **140**, A 1133 (1965)
- [8] R.M. Dreizler and E.K.U. Gross: *Density Functional Theory: An Approach to the Quantum Many-Body Problem* (Springer-Verlag, Berlin, 1990)
- [9] R.O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989)
- [10] R.G. Parr and W. Yang: *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989)
- [11] D.C. Langreth and M.J. Mehl, *Phys. Rev. B* **28**, 1809 (1983)
- [12] J.P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996)
- [13] J.P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **78**, 1396 (1997)
- [14] J. Fritsch and U. Schröder, *Phys. Rep.* **309**, 209 (1999)
- [15] M.T. Yin and M.L. Cohen, *Phys. Rev. Lett.* **45**, 1004 (1980)
- [16] K. Kunc and R.M. Martin, *J. Phys. (Paris)* **42**, C6 (1981)
- [17] K. Kunc and R.M. Martin, *Phys. Rev. Lett.* **48**, 406 (1982)
- [18] M.T. Yin and M.L. Cohen, *Phys. Rev. B* **25**, 4317 (1982)
- [19] W. Frank, C. Elsässer, and M. Fähnle, *Phys. Rev. Lett.* **74**, 1791 (1995)
- [20] G. Kresse, J. Furthmüller, and J. Hafner, *Europhys. Lett.* **32**, 729 (1995)
- [21] K. Parlinski, Z.Q. Li, and Y. Kawazoe, *Phys. Rev. Lett.* **78**, 4063 (1997)

- [22] G.J. Ackland, M.C. Warren, and S.J. Clark, *J. Phys. Condens. Matt.* **9**, 7861 (1997)
- [23] R.P. Feynman, *Phys. Rev.* **56**, 340 (1939)
- [24] M.S. Hybertsen and S.G. Louie, *Phys. Rev. B* **35**, 5585 (1987)
- [25] S.L. Adler, *Phys. Rev.* **126**, 413 (1962)
- [26] N. Wiser, *Phys. Rev.* **129**, 62 (1963)
- [27] R.M. Pick, M.H. Cohen, and R.M. Martin, *Phys. Rev. B* **1**, 910 (1970)
- [28] R. Resta, *Adv. Solid State Phys.* **25**, 183 (1985)
- [29] W.R. Hanke, *Phys. Rev. B* **8**, 4585 (1973)
- [30] N.E. Zein, *Sov. Phys. Solid State* **26**, 1825 (1984)
- [31] D.K. Blat, N.E. Zein, and V.I. Zinenko, *J. Phys. Condens. Matt.* **3**, 5515 (1991)
- [32] N.E. Zein, *Phys. Lett. A* **161**, 526 (1992)
- [33] S. Baroni, P. Giannozzi, and A. Testa, *Phys. Rev. Lett.* **58**, 1861 (1987)
- [34] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni, *Phys. Rev. B* **43**, 7231 (1991)
- [35] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, *Rev. Mod. Phys.* **73**, 515 (2001)
- [36] X. Gonze and J.-P. Vigneron, *Phys. Rev. B* **39**, 13120 (1989)
- [37] X. Gonze, *Phys. Rev. A* **52**, 1086 (1995)
- [38] X. Gonze, *Phys. Rev. A* **52**, 1096 (1995)
- [39] S. de Gironcoli, *Phys. Rev. B* **51**, 6773 (1995)
- [40] A. Dal Corso, *J. Phys.: Condens. Matter* **20**, 445202 (2008)
- [41] M.J. Verstraete, M. Torrent, F. Jollet, G. Zérah, and X. Gonze, *Phys. Rev. B* **78**, 045119 (2008)
- [42] G. Grimvall: *The Electron–Phonon Interaction in Metals* in E. Wohlfarth (ed.): *Selected Topics in Solid State Physics* (North-Holland, New York, 1981)
- [43] R. Heid, K.-P. Bohnen, I.Yu. Sklyadneva, and E.V. Chulkov, *Phys. Rev. B* **81**, 174527 (2010)
- [44] F. Giustino, M.L. Cohen, and S.G. Louie, *Phys. Rev. B* **76**, 165108 (2007)

- [45] M. Calandra, G. Profeta, and F. Mauri, Phys. Rev. B **82**, 165111 (2010)
- [46] P.B. Allen, Phys. Rev. B **6**, 2577 (1972)
- [47] F. Weber, S. Rosenkranz, L. Pintschovius, J.-P. Castellan, R. Osborn, W. Reichardt, R. Heid, K.-P. Bohnen, E.A. Goremychkin, A. Kreyssig, K. Hradil, and D. Abernathy, Phys. Rev. Lett. **109**, 057001 (2012)
- [48] A.M. Saitta, M. Lazzeri, M. Calandra, and F. Mauri, Phys. Rev. Lett. **100**, 226401 (2008)
- [49] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)
- [50] D.J. Scalapino: *The Electron-Phonon Interaction and Strong-Coupling Superconductors* in R.D. Parks (ed.): *Superconductivity*, Vol. 1 (Dekker, New York, 1969)
- [51] G.M. Eliashberg, Zh. Eksp. Fiz. **38**, 966 (1960);
Engl. Transl.: Sov. Phys. – JETP **11**, 696 (1960)
- [52] P.B. Allen and B. Mitrović, Solid State Physics **37**, 1 (1982)
- [53] B.N. Brockhouse, T. Arase, G. Cagliotti, K.R. Rao, and A.D.B. Woods, Phys. Rev. **128**, 1099 (1962)
- [54] W.L. McMillan and J.M. Rowell: *Tunneling and Strong-Coupling Superconductivity* in R.D. Parks (ed.): *Superconductivity*, Vol. 1 (Dekker, New York, 1969)
- [55] J. Geerk, R. Schneider, G. Linker, A.G. Zaitsev, R. Heid, K.-P. Bohnen, and H.v. Löhneysen, Phys. Rev. Lett. **94**, 227005 (2005)
- [56] Ph. Hofmann, I.Yu. Sklyadneva, E.D.L. Rienks, and E.V. Chulkov, New J. Phys. **11**, 125005 (2009)
- [57] I.Yu. Sklyadneva, R. Heid, P.M. Echenique, K.-P. Bohnen, and E.V. Chulkov, Phys. Rev. B **85**, 155115 (2012)
- [58] F. Reinert, B. Eltner, G. Nicolay, D. Ehm, S. Schmidt, and S. Hufner, Phys. Rev. Lett. **91**, 186406 (2003); *ibid* Phys. Rev. Lett. **92**, 089904 (2004) (Erratum)
- [59] A. Eiguren, C. Ambrosch-Draxl, and P.M. Echenique, Phys. Rev. B **79**, 245103 (2009)

13 Eliashberg Theory

Giovanni A.C. Ummarino

Departement of Applied Science and Technology

Politecnico di Torino, 10129 Turin, Italy

Contents

1	Introduction	2
2	Imaginary-axis Eliashberg equations	5
2.1	Nambu formalism	5
2.2	Migdal-Eliashberg theory	6
2.3	Coulomb pseudopotential	10
3	Real-axis Eliashberg equations	11
4	Simplified approaches	12
4.1	BCS limit	12
4.2	Critical temperature	12
5	Relation between real- and imaginary-axis formulation	13
5.1	Padé method for analytic continuation	13
5.2	Marsiglio, Schossmann, and Carbotte formulation	15
6	Tunneling inversion of the standard Eliashberg equations	15
7	Approximations of the standard Eliashberg equations	18
8	Cuprate high-temperature superconductors	19
9	Multi-band Eliashberg theory	21
10	Iron pnictide superconductors	24
10.1	Gaps, critical temperature, and upper critical magnetic field	24
10.2	Interaction mechanism	31
11	Conclusion	32

1 Introduction

The first microscopic theory for superconductivity was proposed in 1957 by Bardeen, Cooper and Schrieffer [1] almost 50 years after the discovery of Kamerlingh Onnes of the zero electric resistance of the mercury below 4.1 K. The theory of superconductivity can be divided into two separate areas: first the establishment of a pairing formalism, which leads to a superconducting condensate, given some attractive particle-particle interaction, and secondly, a mechanism by which two electrons might attract one another. BCS [2], by simplifying the interaction, succeeded in establishing the pairing formalism. Indeed, one of the elegant outcomes of the BCS pairing formalism is the universality of various properties; at the same time this universality means that the theory really does not distinguish one superconductor from another, and, more seriously, one mechanism from another. Luckily, while many superconductors do display universality, some do not, and these, as it turns out, provide very strong support for the electron-phonon mechanism. Before one establishes a theory of superconductivity, one requires a satisfactory theory of the normal state [3]. In conventional superconductors, the Fermi liquid theory appears to work very well, so that, while we cannot solve the problem of electrons interacting through the Coulomb interaction, experiment tells us that Coulomb interactions give rise to well-defined quasiparticles, i.e., a set of excitations which are in one-to-one correspondence with those of the free-electron gas. The net result is that one begins the problem with a reduced Hamiltonian

$$H_{\text{red}} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\mathbf{k}'} V_{\mathbf{k},\mathbf{k}'} c_{\mathbf{k}\uparrow}^\dagger c_{-\mathbf{k}\downarrow}^\dagger c_{-\mathbf{k}'\downarrow} c_{\mathbf{k}'\uparrow}, \quad (1)$$

where, for example, the electron energy dispersion $\varepsilon_{\mathbf{k}}$ already contains much of the effect due to Coulomb interactions. The important point is that well-defined quasiparticles with a well defined energy dispersion near the Fermi surface are assumed to exist, and are summarized by the dispersion $\varepsilon_{\mathbf{k}}$ with a pairing interaction $V_{\mathbf{k},\mathbf{k}'} \equiv V(\mathbf{k}, \mathbf{k}')$. The BCS equations is

$$\Delta_{\mathbf{k}} = - \sum_{\mathbf{k}'} V_{\mathbf{k},\mathbf{k}'} \frac{\Delta_{\mathbf{k}'}}{2E_{\mathbf{k}'}} \tanh \frac{E_{\mathbf{k}'}}{2T}, \quad (2)$$

where

$$E_{\mathbf{k}} = \sqrt{\varepsilon_{\mathbf{k}}^2 + \Delta_{\mathbf{k}}^2} \quad (3)$$

is the quasiparticle energy in the superconducting state, $\Delta_{\mathbf{k}}$ the variational parameter used by BCS, $N(0)$ the normal density of states at the chemical potential E_F , which is set to zero. An additional equation which must be considered together with the gap equation is the number equation,

$$n = 1 - \frac{1}{N(0)} \sum_{\mathbf{k}'} \frac{\varepsilon_{\mathbf{k}'}}{E_{\mathbf{k}'}} \tanh \frac{E_{\mathbf{k}'}}{2T}. \quad (4)$$

Given a pair potential and an electron density, one has to solve these equations to determine the variational parameter $\Delta_{\mathbf{k}}$ and the chemical potential, generally with an iterative numerical method. For conventional superconductors the chemical potential hardly changes from the

normal to the superconducting state, and the variational parameter is much smaller than the chemical potential, with the result that the second equation was usually ignored. BCS then modeled the pairing interaction as a negative (and therefore attractive) constant potential V with a sharp cutoff in momentum space at the Debye energy ω_D

$$V_{\mathbf{k},\mathbf{k}'} \approx -V \theta(\omega_D - |\varepsilon_{\mathbf{k}}|) \theta(\omega_D - |\varepsilon_{\mathbf{k}'}|). \quad (5)$$

Using this potential in the BCS equation (2) with (3), along with a constant density of states assumption over the entire range of integration,

$$\frac{1}{\lambda} = \int_0^{\omega_D} \frac{\tanh(E/2T)}{E} d\varepsilon \quad (6)$$

where $\lambda = N(0)V$. At $T = 0$ K, the integral can be done analytically to give

$$\Delta = 2\omega_D \frac{\exp(-1/\lambda)}{1 - \exp(-1/\lambda)}. \quad (7)$$

Close to the critical temperature, T_c , the BCS equation becomes

$$\frac{1}{\lambda} = \int_0^{\omega_D/2T_c} \frac{\tanh x}{x} dx \quad (8)$$

which cannot be solved in terms of elementary functions for arbitrary coupling strength. Nonetheless, in weak coupling, one obtains

$$T_c = 1.13 \omega_D \exp(-1/\lambda). \quad (9)$$

It is clear that T_c or the zero temperature variational parameter Δ depends on material properties such as the phonon spectrum ω_D , the electronic structure $N(0)$ and the electron-ion coupling strength V . However, it is possible to form various thermodynamic ratios, that turn out to be independent of material parameters. The obvious example from the preceding equations is the ratio $2\Delta/T_c$. In weak coupling (most relevant for conventional superconductors), for example,

$$\frac{2\Delta}{T_c} = 3.53, \quad (10)$$

that is a universal result, independent of the material involved. Many other such ratios can be determined within BCS theory, and the observed deviations from these universal values contributed to the need for an improved formulation of BCS theory.

In the '60s the first discrepancies between the experimental results and the theoretical predictions began to be observed and the BCS theory [2] turned out to be inadequate for superconductors in which the electron-phonon interaction is strong. A primary reason for this is the instantaneous nature of the BCS interaction which does not incorporate enough of the physics of the electron-phonon system. For example, the electron-phonon interaction causes a mass

	$2\Delta/T_c$	$(C_s - C_n)/C_s$	λ	T_C
Al	3.535	1.43	0.43	1.18
Sn	3.705	1.68	2.77	3.75
Pb	4.497	2.77	1.55	7.22
Hg	4.591	2.49	1.62	4.19

Table 1: *Deviations from the universality of BCS theory for some elemental superconductors.*

enhancement of electron states near the Fermi level, seen in the specific heat, and a finite lifetime of electron quasiparticle states. In many material these effects are really, strong and well-defined quasiparticles no longer exists. Nevertheless, Migdal [4] showed that Feynman-Dyson perturbation theory can solve the electron-phonon problem to high accuracy, because the small parameter $\lambda\Omega_D/E_F \approx 10^{-3}$ keeps higher order corrections small.

Table 1 shows the values of the principal quantities for some characteristic elements. They differ more and more from the BCS predictions as the of coupling constant λ increases. According to BCS theory, the expected values are $2\Delta/T_C = 3.53$ and $(C_s - C_n)/C_s = 1.43$. These deviations arise when the interaction between electrons and phonons is strong, while in the weak-coupling approximation the properties of the lattice and the dispersion of phonon curves do not enter directly into the BCS theory.

The prediction of superconducting properties such as the critical temperature or the superconducting energy gap remains one of the outstanding challenges in modern condensed matter theory. Owing to the complex nature of the superconducting state, a quantitative understanding of the pairing mechanism in superconductors requires a very detailed knowledge of the electronic structure, the phonon dispersions, and the interaction between electrons and phonons. For example, in conventional superconductors below the critical temperature electron pairing results from a subtle interplay between the repulsive Coulomb interaction and the attractive electron-phonon interaction. Starting from the BCS theory several approaches to the calculation of the superconducting properties have been proposed for arriving at first-principles Green's function methods such as the Migdal-Eliashberg [5] formalism that provides a very accurate description of the superconducting state in almost all superconductors.

The electron-phonon coupling provided by Eliashberg theory is local in space and retarded in time, reflecting the delay in the development of lattice overscreening. The result is in contrast to the non local, instantaneous nature of the BCS model interaction, attractive for any pair of electrons within the Debye energy ω_D of the Fermi surface. Eliashberg theory is valid only when $\lambda\omega_D/E_F (\simeq \sqrt{m^*/M}) \ll 1$, where E_F is the Fermi level. This is the range of Migdal's theorem.

Migdal [4] argued that all vertex corrections are $\mathcal{O}(\sqrt{m^*/M})$, where m^* is the electron effective mass and M the ion mass, compared to the bare vertex, and therefore can be ignored; this means that only single phonon scattering terms will contribute to the electronic self energy.

2 Imaginary-axis Eliashberg equations

2.1 Nambu formalism

The Fröhlich interaction is formally very similar to the electron-electron interaction via Coulomb forces, thus the mutual scattering of two electrons can be explained through the electron-phonon-electron interaction in the same way. But the phase transition to the superconducting state invalidates the perturbation theory developed for a metal in the normal state. However, in 1960, Nambu showed [6] how the formalism used in the normal state can be rewritten in such a way that the diagrams used to deal with the normal state are applicable also to the superconducting state. The inclusion of Coulomb interactions causes the electron-phonon interaction to be screened and this can constitute a considerable reduction.

In spite of the strong electron-phonon coupling, it remains true that phonon corrections to the electron-phonon vertex are small. In contrast, Coulombic corrections are not necessarily small, but are more or less constant factors, so they can be included in the coupling constant. In the Nambu formalism a 2-component spinor for the electron

$$\psi_{\mathbf{k}} = \begin{pmatrix} c_{\mathbf{k}\uparrow} \\ c_{-\mathbf{k}\downarrow}^\dagger \end{pmatrix}, \quad \psi_{\mathbf{k}}^\dagger = \begin{pmatrix} c_{\mathbf{k}\uparrow}^\dagger & c_{-\mathbf{k}\downarrow} \end{pmatrix} \quad (11)$$

and a bare-phonon field operator

$$\varphi_{q\nu} = b_{q\nu} + b_{-q\nu}^\dagger \quad (12)$$

are defined. The Hamiltonian of an electron-phonon interacting system can be written [3] in terms of ψ and φ . Including Coulomb interactions and it becomes

$$\begin{aligned} H = & \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \psi_{\mathbf{k}}^\dagger \varepsilon_{\mathbf{k}} \psi_{\mathbf{k}}^\dagger \sigma_3 \psi_{\mathbf{k}} + \sum_{q\lambda} \Omega_{q\lambda} b_{q\lambda}^\dagger b_{q\lambda} \sum_{\mathbf{k}\mathbf{k}'\lambda} g_{\mathbf{k},\mathbf{k}'\lambda} \varphi_{\mathbf{k}-\mathbf{k}'\lambda} \psi_{\mathbf{k}}^\dagger \sigma_3 \psi_{\mathbf{k}} \\ & + \frac{1}{2} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4} \langle \mathbf{k}_3\mathbf{k}_4 | V_C | \mathbf{k}_1\mathbf{k}_2 \rangle \left(\psi_{\mathbf{k}_3}^\dagger \sigma_3 \psi_{\mathbf{k}_1} \right) \left(\psi_{\mathbf{k}_4}^\dagger \sigma_3 \psi_{\mathbf{k}_2} \right), \end{aligned} \quad (13)$$

where $\varepsilon_{\mathbf{k}}$ is the one-electron Bloch energy relative to the Fermi level E_F , σ_3 is a Pauli matrix,¹ Ω is the bare phonon energy of wavevector \mathbf{q} and mode ν , $g_{\mathbf{k}\mathbf{k}'\nu}$ are electron phonon matrix elements, and V_C is the Coulomb potential.

Translational invariance of V_C restricts $\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4$ to be either zero or a reciprocal lattice vector \mathbf{K} . The electrons are described in an extended zone scheme and the phonons are described in a reduced zone scheme which is extended periodically throughout \mathbf{q} -space. In order to apply perturbation methods to superconductors the possibility of the existence of Cooper pairs has to be included. This can be done taking the anomalous propagators. Using the

¹The Pauli matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Nambu formalism the Green function [7] becomes

$$G(\mathbf{k}, \tau) = -\langle \mathcal{T} \{ \psi_{\mathbf{k}}(\tau) \psi_{\mathbf{k}}^\dagger(0) \} \rangle, \quad (14)$$

$$D_\lambda(\mathbf{q}, \tau) = -\langle \mathcal{T} \{ \varphi_{\mathbf{q}\lambda}(\tau) \varphi_{\mathbf{k}}^\dagger(0) \} \rangle \quad (15)$$

and the average is over the grand canonical ensemble ($\beta = 1/T$, where T is the temperature)

$$\langle Q \rangle = \frac{\text{Tr } e^{-\beta H} Q}{\text{Tr } e^{-\beta H}}, \quad (16)$$

where the operators evolve in imaginary time and \mathcal{T} is the time-ordering operator. As the matrix operator $\psi_{\mathbf{k}}(\tau) \psi_{\mathbf{k}}^\dagger(0)$ does not conserve the number of particles the definition of a new operator U that adjusts the number of particles is necessary:

$$U = 1 + R^\dagger + R, \quad (17)$$

where R converts a given state in an N -particle system into the corresponding state in the $N + 2$ particle system.

By this definition, the Green function for electrons is a 2×2 matrix, the diagonal elements are the conventional Green functions for spin-up electrons and spin-down holes, while G_{12} and G_{21} describe the pairing properties. It is defined as

$$G(\mathbf{k}, \tau) = - \begin{pmatrix} \langle \mathcal{T} \{ c_{\mathbf{k}\uparrow}(\tau) c_{\mathbf{k}\uparrow}^\dagger(0) \} \rangle & \langle U \mathcal{T} \{ c_{\mathbf{k}\uparrow}(\tau) c_{-\mathbf{k}\downarrow}(0) \} \rangle \\ \langle U \mathcal{T} \{ c_{-\mathbf{k}\downarrow}^\dagger(\tau) c_{\mathbf{k}\uparrow}^\dagger(0) \} \rangle & \langle \mathcal{T} \{ c_{-\mathbf{k}\downarrow}^\dagger(\tau) c_{-\mathbf{k}\downarrow}(0) \} \rangle \end{pmatrix}. \quad (18)$$

The diagonal elements are the ‘normal’ propagators. The off-diagonal elements are Gor’kov’s F and \bar{F} , respectively.

The phonon and electron Green function could be expanded in a Fourier series

$$D_\lambda(\mathbf{q}, \tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} e^{-i\nu_n \tau} D_\lambda(\mathbf{q}, i\nu_n) \quad (19)$$

$$G(\mathbf{k}, \tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} e^{-i\omega_n \tau} G(\mathbf{k}, i\omega_n), \quad (20)$$

where $\nu_n = 2n\pi/\beta$ and $\omega_n = (2n+1)\pi/\beta$ with integer n are the Matsubara frequencies. They are odd multiples of π/β for Fermions while for Bosons they are even.

2.2 Migdal-Eliashberg theory

The basic components of a many-body system are the propagators and the Migdal-Eliashberg theory [8–14] is no exception. The one-electron Green function for the non-interacting system, in momentum space at imaginary frequencies, is given by

$$G_0(\mathbf{k}, i\omega_n) = [i\omega_n \mathbf{1} - \varepsilon_{\mathbf{k}} \sigma_3]^{-1} \quad (21)$$

and for the phonons

$$D_0(\mathbf{q}, i\nu_n) = [M (\omega^2(\mathbf{q}) + \nu_n^2)]^{-1}, \quad (22)$$

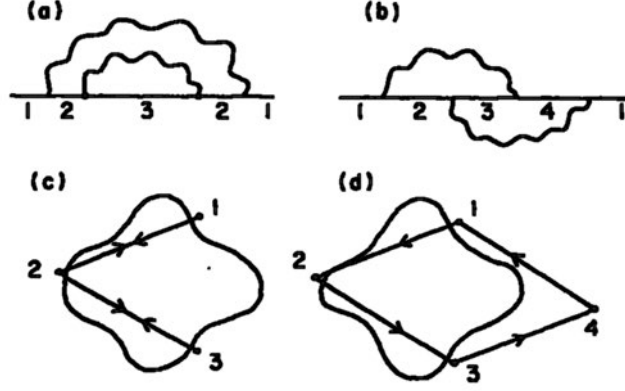


Fig. 1: Feynman diagrams (a) and (b) are the corrections of second order in the electron-phonon interaction to the electron propagator. Diagram (a) is included in Migdal theory while (b) is the first omitted diagram. Panels (c) and (d) shown schematic Fermi surfaces and particular k -states that contribute to (a) and (b), respectively. The last term, in general, will involve large energy denominators (as 3-4 and 1-4) such that it is negligible. This theorem may fail in two circumstances: (i) when either phonon has $|q|$ small, or (ii) when the Fermi surface has a one-dimensional topology.

where M is the ion mass and $\omega(q)$ the phonon dispersion.

From a diagrammatic analysis a Dyson-like equation for the electron and phonon Green functions can be written, though now for the electron it will be a 2×2 matrix equation

$$[G(\mathbf{k}, i\omega_n)]^{-1} = [G_0(\mathbf{k}, i\omega_n)]^{-1} - \Sigma(\mathbf{k}, i\omega_n), \quad (23)$$

$$[D(\mathbf{q}, i\nu_n)]^{-1} = [D_0(\mathbf{q}, i\nu_n)]^{-1} - \Pi(\mathbf{q}, i\nu_n). \quad (24)$$

where Σ is the electronic and Π the phonon self-energy. In principle, in these self-energies, contain the full electron-phonon vertex. Migdal's theorem states that vertex corrections are small. It is therefore a good approximation to set the vertex to the bare vertex, meaning that the electron-phonon interaction is truncated at order $\sqrt{m/M} \sim \omega_D/E_F$. The self energy is then approximated as

$$\Sigma(\mathbf{k}, i\omega_n) = -\frac{1}{\beta} \sum_{\mathbf{k}' n' \nu} \sigma_3 G(\mathbf{k}', i\omega_{n'}) \sigma_3 \left[|g_{\mathbf{k}, \mathbf{k}', \nu}|^2 D_\nu(\mathbf{k} - \mathbf{k}', i\omega_n - i\omega_{n'}) + V_C(\mathbf{k} - \mathbf{k}') \right], \quad (25)$$

where $V_C(\mathbf{k} - \mathbf{k}')$ is the screened Coulomb potential, cf. (13), which has been taken to depend only on the momentum transfer $\mathbf{k} - \mathbf{k}'$.

It is important to remember that $\hat{\Sigma}$ is a 2×2 matrix. It can be rewritten using the Pauli matrices

$$\Sigma(\mathbf{k}, i\omega_n) = i\omega_n [1 - Z(\mathbf{k}, i\omega_n)] \mathbb{1} + \chi(\mathbf{k}, i\omega_n) \sigma_3 + \phi(\mathbf{k}, i\omega_n) \sigma_1 + \bar{\phi}(\mathbf{k}, i\omega_n) \sigma_2. \quad (26)$$

We can now use the Dyson equation to replace the Green function matrix in the Migdal approximation to the self-energy (25). Solving the resulting system of equations for the components of the self-energy will give us the Eliashberg equations.

Using the notation of (26), the electronic Dyson becomes

$$[G(\mathbf{k}, i\omega_n)]^{-1} = i\omega_n Z \mathbb{1} - (\varepsilon_{\mathbf{k}} + \chi) \sigma_3 - \phi \sigma_1 - \bar{\phi} \sigma_2. \quad (27)$$

Inverting this 2×2 matrix, we obtain the Green matrix

$$\begin{aligned} G(\mathbf{k}, i\omega_n) &= \frac{1}{\Theta} \left[i\omega_n Z \mathbf{1} + (\varepsilon_{\mathbf{k}} + \chi) \sigma_3 + \phi \sigma_1 + \bar{\phi} \sigma_2 \right] \\ &= \frac{1}{\Theta} \begin{pmatrix} i\omega_n Z + (\varepsilon_{\mathbf{k}} + \chi) & \phi - i\bar{\phi} \\ \phi + i\bar{\phi} & i\omega_n Z - (\varepsilon_{\mathbf{k}} + \chi) \end{pmatrix}, \end{aligned} \quad (28)$$

with the determinant of (27)

$$\Theta = (i\omega_n Z)^2 - (\varepsilon_{\mathbf{k}} + \chi)^2 - \phi^2 - \bar{\phi}^2. \quad (29)$$

We see that the poles of the Green function matrix, i.e., the electron (and hole) elementary excitations, are given by $\det G(\mathbf{k}, \omega) = \Theta(\mathbf{k}, \omega) = 0$

$$E_{\mathbf{k}} = \sqrt{\left(\frac{\varepsilon_{\mathbf{k}} + \chi}{Z} \right)^2 + \frac{\phi^2 + \bar{\phi}^2}{Z^2}}, \quad (30)$$

thus the gap function is given by

$$\Delta(\mathbf{k}, i\omega_n) = \frac{\phi - i\bar{\phi}}{Z}. \quad (31)$$

Inserting (27) in the Migdal approximation to the self-energy (25), we obtain a system of nonlinear equations for the components of the self-energy. For $\phi = \bar{\phi} = 0$ the system is diagonal and a solution always exists. It corresponds to the normal state. In this case Z and χ are determined by the normal-state self-energy: χ shifts the electronic energies and Z is a renormalization function [8].

If an additional solution with non-zero ϕ or $\bar{\phi}$ exists, it has lower free energy and describes a superconducting state with gap function (31). Actually, it can be shown that, if in the Hamiltonian there are no terms describing spin-dependent interactions, ϕ and $\bar{\phi}$ satisfy identical nonlinear equations hence the solution will have $\phi = \bar{\phi}$, except for a proportionality factor.²

The explicit form of the system of equations for the components of the self-energy is

$$\begin{aligned} i\omega_n [1 - Z(\mathbf{k}, i\omega_n)] &= \frac{1}{\beta} \sum_{\mathbf{k}'n'\nu} |g_{\mathbf{k},\mathbf{k}',\nu}|^2 D_{\nu}(\mathbf{k} - \mathbf{k}', i\omega_n - i\omega_{n'}) \frac{i\omega_{n'} Z(\mathbf{k}', i\omega_{n'})}{\Theta(\mathbf{k}', i\omega_{n'})} \\ \chi(\mathbf{k}, i\omega_n) &= \frac{1}{\beta} \sum_{\mathbf{k}'n'\nu} |g_{\mathbf{k},\mathbf{k}',\nu}|^2 D_{\nu}(\mathbf{k} - \mathbf{k}', i\omega_n - i\omega_{n'}) \frac{\chi(\mathbf{k}', i\omega_{n'}) + \varepsilon'_{\mathbf{k}}}{\Theta(\mathbf{k}', i\omega_{n'})} \\ \phi(\mathbf{k}, i\omega_n) &= -\frac{1}{\beta} \sum_{\mathbf{k}'n'\nu} [|g_{\mathbf{k},\mathbf{k}',\nu}|^2 D_{\nu}(\mathbf{k} - \mathbf{k}', i\omega_n - i\omega_{n'}) - V_C(\mathbf{k} - \mathbf{k}')] \frac{\phi(\mathbf{k}', i\omega_{n'})}{\Theta(\mathbf{k}', i\omega_{n'})} \\ \bar{\phi}(\mathbf{k}, i\omega_n) &= -\frac{1}{\beta} \sum_{\mathbf{k}'n'\nu} [|g_{\mathbf{k},\mathbf{k}',\nu}|^2 D_{\nu}(\mathbf{k} - \mathbf{k}', i\omega_n - i\omega_{n'}) - V_C(\mathbf{k} - \mathbf{k}')] \frac{\bar{\phi}(\mathbf{k}', i\omega_{n'})}{\Theta(\mathbf{k}', i\omega_{n'})} \\ n &= 1 - \frac{2}{\beta} \sum_{\mathbf{k}'n'} \frac{\chi(\mathbf{k}', i\omega_{n'}) + \varepsilon_{\mathbf{k}'}}{\Theta(\mathbf{k}', i\omega_{n'})} \end{aligned}$$

²The arbitrary phase comes from the one of the one-electron state. Normally, the physical quantities cannot depend on this phase. However, it is measured by Josephson tunnelling. Thus BCS theory exhibits a broken gauge symmetry.

These are the *Eliashberg equations*. The last equation gives the electron density and determines the chemical potential μ .

The general Eliashberg equations couple all momenta \mathbf{k} . To simplify them, they are usually averaged over by replacing the sums over momenta by integrals over energy, weighted with the density of states. The result is a single set of equations. This approximation turns out to be good for elemental superconductors, but fails in describing more complex systems.

The \mathbf{k} -dependence in G comes mainly from the explicit $\varepsilon_{\mathbf{k}}$ dependence of Θ , while it can be averaged out in Z and ϕ (fixing $\varepsilon_{\mathbf{k}} = E_F$ because these quantities are non-zero only near the Fermi surface), so

$$\begin{aligned} Z(\mathbf{k}, i\omega_n) &\rightarrow \langle Z(\mathbf{k}, i\omega_n) \rangle_{\varepsilon=E_F} = Z(i\omega_n) \\ \phi(\mathbf{k}, i\omega_n) &\rightarrow \langle \phi(\mathbf{k}, i\omega_n) \rangle_{\varepsilon=E_F} = \phi(i\omega_n) \\ \chi(\mathbf{k}, i\omega_n) &\rightarrow \langle \chi(\mathbf{k}, i\omega_n) \rangle_{\varepsilon=E_F} = \chi(i\omega_n) \end{aligned}$$

On the right hand side of the Eliashberg equations the same \mathbf{k} average can be obtained by applying the operator $\frac{1}{N(0)} \sum_{\mathbf{k}} \delta(\varepsilon_{\mathbf{k}})$ where $N(0)$ is the normal density of states at the Fermi level and introducing a unity factor $\int d\omega \delta(\omega - \omega_{\mathbf{q},\nu})$, where $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ is the phonon wavevector³

$$\begin{aligned} [1 - Z(i\omega_n)] i\omega_n &= -\frac{1}{\beta N^2(0)} \sum_{n'} \int d\omega \sum_{\mathbf{k}\mathbf{k}'\nu} \frac{|g_{\mathbf{k},\mathbf{k}',\nu}|^2 \delta(\varepsilon_{\mathbf{k}'}') \delta(\varepsilon_{\mathbf{k}}) \delta(\omega - \omega_{\mathbf{q},\nu}) 2\omega_{\mathbf{q},\nu}}{(\omega_n - \omega_{n'})^2 + \omega_{\mathbf{q},\nu}^2} \\ &\quad \times \int_{-\infty}^{\infty} d\varepsilon \frac{N(\varepsilon) i\omega_{n'} Z(i\omega_{n'})}{\Theta(\varepsilon, i\omega_{n'})} \\ \phi(i\omega_n) &= \frac{1}{\beta N^2(0)} \sum_{n'} \int d\omega \sum_{\mathbf{k}\mathbf{k}'\nu} \frac{|g_{\mathbf{k},\mathbf{k}',\nu}|^2 \delta(\varepsilon_{\mathbf{k}'}') \delta(\varepsilon_{\mathbf{k}}) \delta(\omega - \omega_{\mathbf{q},\nu}) 2\omega_{\mathbf{q},\nu}}{(\omega_n - \omega_{n'})^2 + \omega_{\mathbf{q},\nu}^2} \\ &\quad \times \int_{-\infty}^{\infty} d\varepsilon \frac{N(\varepsilon) \phi(i\omega_{n'})}{\Theta(\varepsilon, i\omega_{n'})} \\ \chi(i\omega_n) &= -\frac{1}{\beta N^2(0)} \sum_{n'} \int d\omega \sum_{\mathbf{k}\mathbf{k}'\nu} \frac{|g_{\mathbf{k},\mathbf{k}',\nu}|^2 \delta(\varepsilon_{\mathbf{k}'}') \delta(\varepsilon_{\mathbf{k}}) \delta(\omega - \omega_{\mathbf{q},\nu}) 2\omega_{\mathbf{q},\nu}}{(\omega_n - \omega_{n'})^2 + \omega_{\mathbf{q},\nu}^2} \\ &\quad \times \int_{-\infty}^{\infty} d\varepsilon \frac{N(\varepsilon) [\varepsilon + \chi(i\omega_{n'})]}{\Theta(\varepsilon, i\omega_{n'})} \\ n &= 1 - \frac{2}{\beta N(0)} \sum_{n'} \int_{-\infty}^{\infty} d\varepsilon \frac{N(\varepsilon) [\varepsilon + \chi(i\omega_{n'})]}{\Theta(\varepsilon, i\omega_{n'})}. \end{aligned}$$

as the phonon interaction is very low, the sum over \mathbf{k}' has been split up into an angular average for $\varepsilon_{\mathbf{k}} = E_F$ and an integration in ε on the ε dependence of the electronic Green function. Only the states near the Fermi level will contribute to this integral, because of the $\varepsilon_{\mathbf{k}}$ terms in $\Theta(\varepsilon, i\omega_n)$. When the density of states can be considered constant in this region, a further simplification can be introduced using $N(0)$ instead of $N(\varepsilon)$ and performing the integrals analytically. In this way the final result is $\chi(i\omega_n) = 0$ and $n = 1$ (half-filling approximation).

³The equation for χ will be omitted from now on because, in many cases, its contribution can be neglected.

It is useful to define the electron-boson spectral function, the positive-definite function

$$\begin{aligned}\alpha^2 F(\omega) &= N(0) \sum_{q\nu} g_{q,\nu}^2 \delta(\omega - \omega_{q,\nu}) \\ &= \frac{1}{N(0)} \sum_{\mathbf{k}\mathbf{k}'} \sum_{\nu} |g_{\mathbf{k},\mathbf{k}',\nu}|^2 \delta(\varepsilon_{\mathbf{k}'}) \delta(\varepsilon_{\mathbf{k}}) \delta(\omega - \omega_{q,\nu}),\end{aligned}\quad (32)$$

where

$$g_{q,\nu}^2 = \frac{1}{N^2(0)} \sum_{\mathbf{k}'} |g_{\mathbf{k},\mathbf{k}',\nu}|^2 \delta(\varepsilon_{\mathbf{k}+\mathbf{q}}) \delta(\varepsilon_{\mathbf{k}}) \quad (33)$$

is the q -dependent electron-phonon coupling. With this the Eliashberg system takes the form

$$\begin{aligned}[1 - Z(i\omega_n)] i\omega_n &= -\frac{\pi}{\beta} \sum_{n'} \frac{Z(i\omega_{n'}) i\omega_{n'}}{\Xi(i\omega_{n'})} \int d\Omega \frac{2\Omega \alpha^2 F(\Omega)}{(\omega_n - \omega_{n'})^2 + \Omega^2} \\ \phi(i\omega_n) &= \frac{\pi}{\beta} \sum_{n'} \frac{\phi(i\omega_{n'})}{\Xi(i\omega_{n'})} \left[\int d\Omega \frac{2\Omega \alpha^2 F(\Omega)}{(\omega_n - \omega_{n'})^2 + \Omega^2} - N(0) \bar{V}_C \right] \\ \Xi(i\omega_n) &= \sqrt{[Z(i\omega_n) \omega_n]^2 + [\phi(i\omega_n)]^2},\end{aligned}$$

where \bar{V}_C represents an appropriate Fermi surface average of the screened Coulomb potential V_C . The sum over Matsubara frequencies can be cut off at an energy ω_C . Solving these equations, we obtain the electron self-energy at the Fermi level.

2.3 Coulomb pseudopotential

Including the repulsive term in the Eliashberg equations is a hard task. The Coulomb interaction cannot be introduced with the same accuracy of the electron-phonon one, since it does not have a natural cut-off to ensure a convergent sum over the Matsubara frequencies. While the electron-electron interaction has a large energy scale and a correspondingly short interaction time, the electron-phonon interaction has a timescale typical of the much larger inverse phonon frequencies. The time scale difference is normally dealt using an energy window ω_C with a renormalized electron-electron interaction [13]

$$\mu^* = \frac{\mu}{1 + \mu \ln(E_F/\omega_C)}, \quad (34)$$

which is called *Morel-Anderson pseudopotential*. In this formula, μ is an average electron-electron matrix element times the density of states at the Fermi level.

In the normal state self-energy the Coulomb potential is already included, so that only the off-diagonal term will be affected by this correction, giving

$$\phi_C(i\omega_n) = -\mu^* \frac{\pi}{\beta} \sum_{n'} \frac{\phi(i\omega_{n'})}{\Xi(i\omega_{n'})} \theta(\omega_C - |\omega_{n'}|). \quad (35)$$

Including this contribution in the Eliashberg equation for ϕ , we obtain

$$\begin{aligned}\Delta(i\omega_n) Z(i\omega_n) &= \frac{\pi}{\beta} \sum_{n'} \frac{\Delta(i\omega_{n'})}{\sqrt{\omega_{n'}^2 + \Delta^2(i\omega_n)}} [\lambda(i\omega_{n'} - i\omega_n) - \mu^*(\omega_C)] \theta(\omega_C - |\omega_{n'}|) \\ Z(i\omega_n) &= 1 + \frac{\pi}{\omega_n \beta} \sum_{n'} \frac{\omega_{n'}}{\sqrt{\omega_{n'}^2 + \Delta^2(i\omega_n)}} \lambda(i\omega_{n'} - i\omega_n)\end{aligned}$$

where $\lambda(i\omega_{n'} - i\omega_n)$ is related to the electron-boson spectral density $\alpha^2 F(\omega)$ through the relation

$$\lambda(i\omega_{n'} - i\omega_n) = \int_0^\infty d\Omega \frac{2\Omega \alpha^2 F(\Omega)}{\Omega^2 + (\omega_{n'} - \omega_n)^2}. \quad (36)$$

3 Real-axis Eliashberg equations

The Green function can be analytically continued onto the real-frequencies axis, by using the expression $\omega + i\delta$, with an infinitesimal δ . The density of states is contained in the imaginary part of $G(\mathbf{k}, \omega + i\delta)$.

In their real-axis formulation, the Eliashberg equations are a set of two non-linear integral equations for a complex frequency-dependent gap $\Delta(\omega)$ and a renormalization function $Z(\omega)$, which exists also in the normal state. Both $\Delta(\omega)$ and $Z(\omega)$ are temperature dependent.

$$\begin{aligned} \Delta(\omega, T) Z(\omega, T) &= \int_0^{\omega_C} d\omega' \Re \left[\frac{\Delta(\omega', T)}{\sqrt{\omega'^2 - \Delta^2(\omega', T)}} \right] \int_0^\infty d\Omega \alpha^2 F(\Omega) \\ &\quad \times \left\{ [n(\Omega) + f(-\omega')] \left[\frac{1}{\omega + \omega' + \Omega + i\delta^+} - \frac{1}{\omega - \omega' - \Omega + i\delta^+} \right] \right. \\ &\quad \left. - [n(\Omega) + f(\omega')] \left[\frac{1}{\omega - \omega' + \Omega + i\delta^+} - \frac{1}{\omega + \omega' - \Omega + i\delta^+} \right] \right\} \\ &\quad - \mu^* \int_0^{\omega_C} d\omega' \Re \left[\frac{\Delta(\omega', T)}{\sqrt{\omega'^2 - \Delta^2(\omega', T)}} \right] [1 - 2f(\omega')] \\ [1 - Z(\omega, T)] \omega &= \int_0^\infty d\omega' \Re \left[\frac{\omega'}{\sqrt{\omega'^2 - \Delta^2(\omega', T)}} \right] \int_0^\infty d\Omega \alpha^2 F(\Omega) \\ &\quad \times \left\{ [n(\Omega) + f(-\omega')] \left[\frac{1}{\omega + \omega' + \Omega + i\delta^+} - \frac{1}{\omega - \omega' - \Omega + i\delta^+} \right] \right. \\ &\quad \left. - [n(\Omega) + f(\omega')] \left[\frac{1}{\omega - \omega' + \Omega + i\delta^+} - \frac{1}{\omega + \omega' - \Omega + i\delta^+} \right] \right\}. \end{aligned}$$

Here, ω_C is the boson energy cut-off introduced in the Coulomb pseudo potential and $f(\omega) = 1/(e^{\beta\omega} + 1)$ is the Fermi, $n(\Omega) = 1/(e^{\beta\Omega} - 1)$ the Bose function. The real part of the product $\Delta(\omega, T)Z(\omega, T)$ and of $Z(\omega, T)$ is determined by the principal-value integrals, while the imaginary part comes from the delta-function parts.

The denominators can vanish for particular energies. Then the integrals must be done carefully when a numerical approach is used. The low frequency behaviour of the various functions is, at $T = 0$, $\Re[\Delta(\omega)] = c$, $\Im[\Delta(\omega)] = 0$, $\Re[Z(\omega)] = d$ and $\Im[Z(\omega)] = 0$ while, at $T \neq 0$, $\Re[\Delta(\omega)] \propto \omega^2$, $\Im[\Delta(\omega)] \propto \omega$, $\Re[Z(\omega)] = d(T)$ and $\Im[Z(\omega)] \propto 1/\omega$ where c and d are constants.

4 Simplified approaches

4.1 BCS limit

In order to better understand the Eliashberg equations, it can be useful to reduce them to BCS limit. To achieve this aim further approximations are introduced: First of all, the bosons factor in the real-axis Eliashberg equations are ignored, i.e., real bosons scattering are not taken into account. Further, the imaginary parts of Δ and Z must be neglected. We set

$$\Delta(\omega, T) = \begin{cases} \Delta_0(T) & \text{for } \omega < \omega_D \\ 0 & \text{for } \omega \geq \omega_D \end{cases}, \quad (37)$$

where $\Delta_0(T)$ is a real number and ω_D is the Deybe energy and replace $Z(\omega, T)$ by its value in the normal state at $\omega = 0$ and $T = 0$. Then

$$Z(0, T) - 1 = 2 \int_0^\infty d\omega' \int_0^\infty d\Omega \alpha^2 F(\Omega) \left[\frac{f(-\omega')}{(\omega' + \Omega)^2} + \frac{f(\omega')}{(\omega' + \Omega)^2} \right] \equiv \lambda(T), \quad (38)$$

which in the $T \rightarrow 0$ limit is

$$Z(0, 0) - 1 = \int_0^\infty d\Omega \alpha^2 F(\Omega) \int_0^\infty \frac{2d\omega'}{(\omega' + \Omega)^2} \equiv \lambda. \quad (39)$$

The gap equation becomes

$$\Delta_0(T) = \int_{\Delta_0(T)}^{\omega_D} d\omega' \frac{\Delta_0(T)}{\sqrt{\omega'^2 - \Delta_0^2(T)}} \frac{\lambda - \mu^*}{1 + \lambda} [1 - 2f(\omega')]. \quad (40)$$

It is interesting to note that now ω_D is important for both the λ and the μ^* contribution.

With $\varepsilon = \sqrt{\omega'^2 - \Delta_0^2}$, the equation can be rewritten as

$$\Delta_0(T) = \frac{\lambda - \mu^*}{1 + \lambda} \int_0^{\omega_D} d\varepsilon \frac{\Delta_0(T)}{\sqrt{\varepsilon^2 + \Delta_0^2(T)}} [1 - 2f(\sqrt{\varepsilon^2 + \Delta_0^2(T)})], \quad (41)$$

which is the usual BCS equation at finite temperature. In the the $T \rightarrow 0$ limit it becomes

$$\Delta_0 = \frac{\lambda - \mu^*}{1 + \lambda} \int_0^{\omega_D} d\varepsilon \frac{\Delta_0}{\sqrt{\varepsilon^2 + \Delta_0^2}}, \quad (42)$$

which corresponds to the BCS gap equation if we define $\lambda_{\text{BCS}} = (\lambda - \mu^*)/(1 + \lambda)$. The renormalization factor $1/(1 + \lambda)$ comes from the Z term in the Eliashberg equation, i.e., from having included electron-phonon effects.

4.2 Critical temperature

Solving the Eliashberg system, even in the isotropic form, is a quite demanding task. However the most relevant results can be obtained using a simpler approach proposed by McMillan [15]. Through a fit of a large set of results obtained using the spectral function of lead and solving the

Eliashberg equations in a small range of the parameter space ($\lambda < 2$ and $\mu^* < 0.15$), McMillan obtained an analytic formula for the critical temperature:

$$T_C = \frac{\Theta_D}{1.45} \exp \left[-\frac{1.04(1 + \lambda)}{\lambda - \mu^*(1 + 0.62\lambda)} \right], \quad (43)$$

where Θ_D is the Debye temperature and the number λ has the same meaning as the electron-phonon coupling parameter, and can be derived from the Eliashberg function as

$$\lambda = 2 \int d\Omega \frac{\alpha^2 F(\Omega)}{\Omega}. \quad (44)$$

Later, this formula was refined by Allen and Dynes [15], who substituted the factor $\Theta_D/1.45$ with $\Omega_{log}/1.2$, with the much more representative frequency

$$\Omega_{log} = \exp \left[\frac{2}{\lambda} \int d\Omega \log \Omega \frac{\alpha^2 F(\Omega)}{\Omega} \right], \quad (45)$$

which is a weighted average of the phonon frequencies. The McMillan formula predicts an upper limit for T_C even if λ increases indefinitely. However this was a wrong conclusion because the equation (43) was not derived analytically but obtained by numerical solutions in a fixed range of the coupling constant and then it is not possible to consider the limit for $\lambda \rightarrow \infty$. For $\lambda \gg 1$, taking the limit of the Eliashberg equations the following expression for T_C can be obtained in an analytical way

$$T_C = 0.183\omega_D\sqrt{\lambda} \quad (46)$$

and it is clear that in Eliashberg theory there is no upper limit for the critical temperature.

In general the Eliashberg equations are solved numerically with an iterative method until you reach self-consistency. The numerical procedure is very simple in the formulation on imaginary axis, much less so on the real one. The critical temperature can be calculated or by solving an eigenvalue equation [8] or, more easily, by giving a very small test value to superconducting gap (for the Pb it is $\Delta = 1.4$ meV at $T = 0$ K so, for example, $\Delta(T) = 10^{-7}$ meV) and checking at which temperature the solution converges. In this way, T_c is obtained with accuracy superior to experimental error bars.

5 Relation between real- and imaginary-axis formulation

5.1 Padé method for analytic continuation

The Eliashberg equations on the real axis are very difficult to solve, while their formulation on the imaginary axis, while simpler to solve, can be used almost only to evaluate the critical temperature. Therefore, a procedure which allows obtaining the real-axis gap and the renormalization function by analytically continuing $\Delta(i\omega_n)$ and $Z(i\omega_n)$ is often used [16]. This procedure makes use of Padé approximants. It speed up the numerical solution of Eliashberg equations. However the Padé method is valid only at $T < T_C/10$, thus it is often necessary

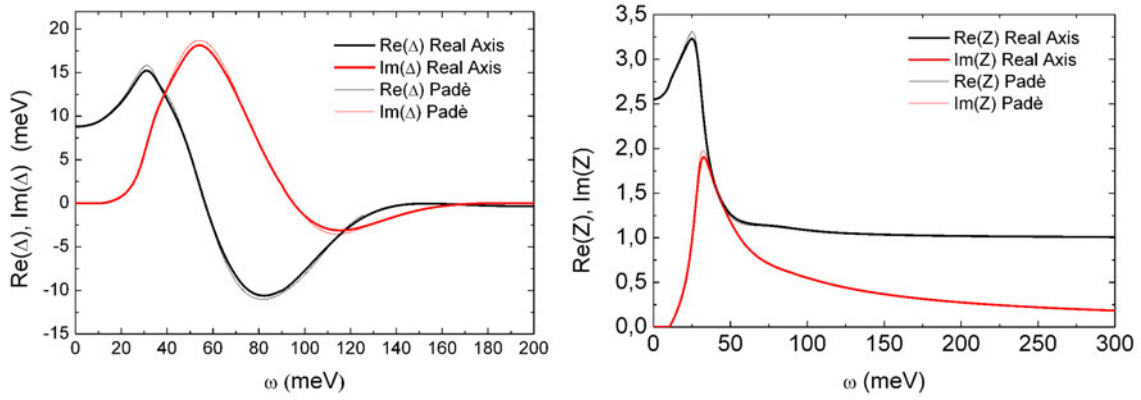


Fig. 2: Validity check of Padé approximants methods. Here real and imaginary part of $\Delta(\omega)$ and $Z(\omega)$ are shown at $T < T_C/10$.

to still solve for $\Delta(\omega)$ directly from the real-frequency equations. Also when the Eliashberg equations contain some terms that describe the presence of impurities in the superconductor the accuracy of the Padé approximants can leave something to be desired.

The N -point Padé approximant to a complex function $u(z)$ of the complex variable z , whose N values $u_i = u_i(z_i)$ ($i = 1, \dots, N$) are given at N complex points z_i , is defined as the continued fraction

$$C_N(z) = \frac{a_1}{1 + \frac{a_2(z - z_1)}{1 + \frac{a_3(z - z_2)}{1 + \frac{a_4(z - z_3)}{\ddots \frac{1}{1 + a_n(z - z_{n-1})}}}}} \quad (47)$$

such that

$$C_N(z_i) = u_i, \quad i = 1, \dots, N. \quad (48)$$

The coefficients a_i are given by recursive formula $a_i = g_i(z_i)$, where

$$\begin{aligned} g_1(z_i) &= u_i && \text{with } i = 1, \dots, N \\ g_p(z) &= \frac{g_{p-1}(z_{p-1}) - g_{p-1}(z)}{(z - z_{p-1})g_{p-1}(z)} && \text{for } p \geq 2 \end{aligned}$$

It can be shown that the continued fraction can be evaluated order-by-order via $C_N(z) = A_N(z)/B_N(z)$ where A_N and B_N are polynomials given by the recursion relation

$$\begin{aligned} A_{n+1}(z) &= A_n(z) + (z - z_n)a_{n+1}A_{n-1}(z) && \text{with } n = 1, 2, \dots, N-1 \\ B_{n+1}(z) &= B_n(z) + (z - z_n)a_{n+1}B_{n-1}(z) && \text{with } n = 1, 2, \dots, N-1 \end{aligned}$$

with starting values $A_0 = 0$, $A_1 = a_1$, and $B_0 = B_1 = 1$.

A comparison of results obtained with the real-axis equations and results obtained with the Padé method is shown in Fig. 2.

5.2 Marsiglio, Schossmann, and Carbotte formulation

A more recent method [10] for solving the Eliashberg equations on the real axis introduces two equations for the renormalized frequency $\tilde{\omega}(z) = zZ(z)$ and the pairing function $\phi(z)$

$$\begin{aligned} \tilde{\omega}(\omega) = & \omega + i\pi T \sum_{m=1}^{\infty} \frac{\tilde{\omega}(i\omega_m)}{\sqrt{\tilde{\omega}^2(i\omega_m) + \phi^2(i\omega_m)}} [\lambda(\omega - i\omega_m) - \lambda(\omega + i\omega_m)] \\ & + i\pi \int_{-\infty}^{\infty} dz \frac{\tilde{\omega}(\omega - z)}{\sqrt{\tilde{\omega}^2(\omega - z) - \phi^2(\omega - z)}} \alpha^2 F(z) [n(z) + f(z - \omega)] \end{aligned} \quad (49)$$

$$\begin{aligned} \phi(\omega) = & i\pi T \sum_{m=1}^{\infty} \frac{\phi(i\omega_m)}{\sqrt{\tilde{\omega}^2(i\omega_m) + \phi^2(i\omega_m)}} [\lambda(\omega - i\omega_m) - \lambda(\omega + i\omega_m) - 2\mu^* \theta(\omega_C - |\omega_m|)] \\ & + i\pi \int_{-\infty}^{\infty} dz \frac{\phi(\omega - z)}{\sqrt{\tilde{\omega}^2(\omega - z) - \phi^2(\omega - z)}} \alpha^2 F(z) [n(z) + f(z - \omega)]. \end{aligned} \quad (50)$$

These equations give solutions for the real-axis gap and renormalization function that are identical to those obtained from the solution of the real-axis equations. Then they are valid at any temperature, but their numerical solution presents problems completely analogous to those of the formulation on the real axis. The choice between these equations and those on the real axis is thus just a matter of personal taste.

6 Tunneling inversion of the standard Eliashberg equations

In the past $\alpha^2 F(\Omega)$ and μ^* were obtained by experimental data or were considered as free parameters. As shown in the lecture of R. Heid, now they can actually be calculated by density functional theory [17] so, the theory does not contain free parameters.

When such calculations are not possible, the standard method to obtain these physical input parameters involves single-particle tunneling spectroscopy, which is perhaps the simplest, most direct probe of the excitations of a solid. In these experiments electrons are injected into (or extracted from) a sample, as a function of bias voltage V . The typical example is a planar junction SIN (superconductor, thin layer of insulating and a metal in the normal state) [18]. The resulting current is proportional to the superconducting density of states [19]

$$I_S(V) \propto \int d\omega \Re \left[\frac{|\omega|}{\sqrt{\omega^2 - \Delta^2(\omega)}} \right] [f(\omega) - f(\omega + V)], \quad (51)$$

where we have used the gap function, $\Delta(\omega)$, defined as $\Delta(\omega) \equiv \phi(\omega)/Z(\omega)$. The proportionality constant contains information about the density of states of the normal metal and the tunneling matrix element. They are usually assumed to be constant.

In the zero temperature limit the derivative of the current with respect to the voltage is simply proportional to the superconducting density of states

$$\left(\frac{dI}{dV} \right)_S / \left(\frac{dI}{dV} \right)_N \propto \Re \left\{ \frac{|V|}{\sqrt{V^2 - \Delta^2(V)}} \right\}, \quad (52)$$

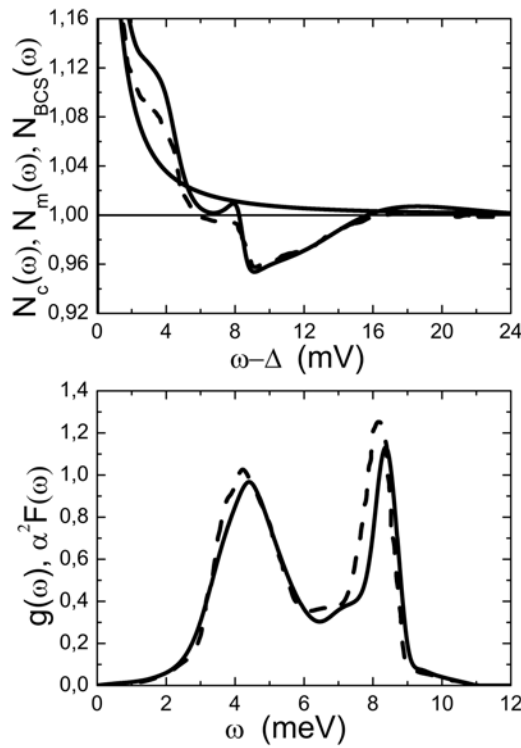


Fig. 3: Upper panel: Energy dependence of the lead density of states in the superconducting states at very low temperature ($T \ll T_c$). The dashed line represents experimental data obtained from tunneling, while BCS weak-coupling theory yields the dash-dotted line. The Eliashberg strong coupling result is shown by a solid line. Lower panel: Calculated electron phonon spectral function of Pb (solid line) determined by inversion of Eliashberg equations compared with the measured phonon density of states (dashed line) for the same material.

where S and N denote superconducting and normal state, respectively. The right hand side is simply the density of states, computed within the Eliashberg framework. From the experimentally measured quasiparticle density of states at very low temperature $N_m(V) = (\frac{dI}{dV})_S / (\frac{dI}{dV})_N$, where the subscript m denotes "measured", it is possible, through a complicated mathematical procedure, to obtain the electron-phonon spectral function $\alpha^2 F(\omega)$ and the Coulomb pseudopotential μ^* , not only for superconductors but also for normal metals via the proximity effect [19].

The procedure followed is conceptually simple but mathematically involved [20, 21]. A first guess is made for the two quantities, namely $\alpha^2 F_0(\omega)$, i.e., starting with a generic function greater than zero in a finite range and with a Coulomb pseudopotential parameter, $\mu_0^* \simeq 0.1$. So the Eliashberg equations (at $T = 0$) can be solved numerically with these two input parameters in order to obtain the complex function $\Delta(\omega)$ necessary for calculating the superconductive density of states $N_c^0(\omega)$ denoted by the subscript c (calculated) and 0 (for a first choice). Next, the functional derivative $\frac{\delta N_c^0(\omega)}{\delta \alpha^2 F(\nu)}$ which give the infinitesimal response of $N_c^0(\omega)$ to the change in $\alpha^2 F(\nu)$ is computed. This is used to make a second guess for $\alpha^2 F(\nu)$ through the equation $\delta \alpha^2 F_0(\nu) = \int d\omega [\frac{\delta N_c^0(\omega)}{\delta \alpha^2 F(\nu)}]^{-1} [N_m(\omega) - N_c^0(\omega)]$. The new electron phonon spectral function is

$\alpha^2 F_1(\nu) = \alpha^2 F_0(\nu) + \delta \alpha^2 F_0(\nu)$. This procedure is continued until convergence is reached. Unique $\alpha^2 F(\nu)$ and μ^* result. They are referred to as the measured microscopic parameters for that particular material. It is not at all apparent that the structures of the density of states, as can be seen in the upper panel of Fig. 3, reflect the shape of the electron-phonon spectral function, through the function $\Delta(\omega)$.

At zero temperature the gap function $\Delta(\omega)$ is real and roughly constant up to a certain energy equal to that constant. This implies that the density of states will have a gap, as in BCS theory. At finite temperature the gap function has a small imaginary part starting from zero frequency (and, in fact the real part approaches zero at zero frequency so that in principle there is no gap, even for an s-wave order parameter. In practice, a very well-defined gap still occurs for moderate coupling, and disappears at finite temperature only when the coupling strength is increased significantly.

McMillan and Rowell were able to deconvolve their measurement, to produce the single electron density of states. Since the superconducting density of states is given by the right hand side of (52), the structure in the data must reflect the structures in the gap function, $\Delta(\omega)$, that originate from the input function, $\alpha^2 F(\omega)$. In other words, the dI/dV curve can be viewed as as a highly nonlinear transform of $\alpha^2 F(\omega)$. Thus the, usually very small, structure present in the density of states contains important information (in coded form) concerning the electron-phonon interaction as can be seen in the lower panel of Fig. 3, where the calculated electron phonon spectral function of Pb determined by inversion of the Eliashberg equations is compared with the neutron scattering measurements of the phonon density of states for the same material: this is the most clear way to determine the mechanism of superconductivity in a material.

Once $\alpha^2 F(\omega)$ (and μ^*) has been acquired in this way one can use the Eliashberg equations to calculate other properties, for example, T_c or many others physical quantities (temperature dependence of the gap, of the upper critical field, of the specific heat etc). These can then be compared to experiment, and the agreement in general tends to be fairly good.

One may suspect, however, a circular argumentation, since the theory was used to produce the spectrum (from experiment), and now the theory is used as a predictive tool, with the same spectrum. There are a number of reasons, however, for believing that this procedure has produced meaningful information. First of all, the spectrum obtained comes out to be positive definite, as is physically required. Second, the spectrum is non-zero precisely in the phonon region, as it should be and it agrees very well with the measured spectrum. Moreover, as already mentioned, various thermodynamic properties are calculated with this spectrum, with good agreement with experiments. Finally, the density of states itself can be calculated in a frequency regime beyond the phonon region. The agreement with experiment is spectacular.

None of these indicators of success can be taken as definitive proof of the electron-phonon interaction. For example, even the excellent agreement with the density of states could be understood as a mathematical property of analytic functions. This procedure has not been so straightforward or possible in all superconductors.

An alternate inversion procedure, which utilizes a Kramers-Kronig relation to extract $\Delta(\omega)$ from the tunneling result and remove μ^* from the procedure has been provided [22]. At last a

procedure exists for obtaining the electron-phonon spectral density by inversion of optical conductivity data, a process very similar in spirit to the McMillan-Rowell inversion of tunnelling data. This procedure has the advantage that it can be utilized also in the normal state [23].

7 Approximations of the standard Eliashberg equations

As mentioned before, the standard Eliashberg theory has been formulated within a lot of approximations. Here a list of these simplifications with possible generalizations:

- **validity of Migdal's theorem:** In almost all superconductors the condition $\omega_D/E_F \ll 1$ is fulfilled. In HTCS and fullerenes $\omega_D/E_F \sim 10^{-1}$ and it is necessary to include vertex corrections in the self-energy [24].
- **single conduction band:** Before the discovery of MgB_2 the known superconductors could be described within one-band models. Then the theory has been generalized to two (MgB_2) [25] or more bands (iron pnictides).
- **isotropic order parameter:** In the oldest superconductors the order parameter does not depend on the position on the Fermi surface, i.e., $\Delta \equiv \Delta(\mathbf{k})$. There is experimental evidence that this is not true in HTCS [26–28].
- **singlet superconductivity:** Usually the spin of Cooper pairs is equal to zero, but in Sr_2RuO_4 [29] probably it equals one, implying a different spatial symmetry (*p*-wave) [8].
- **infinite conduction bandwidth:** In almost all superconductors the width of the conduction band is much larger than the representative energy of the boson mediating the Cooper pairs interaction (phonons, antiferromagnetic spin fluctuations) so that it can be considered to be infinite. In HTCS and Fullerenes this approximation breaks down and the real bandwidth has to be included in the theory [13].
- **half filling:** Typically the occupation of the conduction band is symmetric. In HTCS this is not true and the number of the Eliashberg equation increases because of $\chi(\omega) \neq 0$ [13].
- **flat normal density of states:** Generally the normal density of states can be approximated by a constant around the Fermi level. In PuCoGa_5 and in a small number of other compounds this approximation is not valid [30]. Also in this case the number of Eliashberg equations increases [8].
- **no disorder or magnetic impurities:** A material can be disordered with chemical doping or neutron irradiation, moreover magnetic impurities can be added. To describe these physical situations, new terms have to be introduced in the Eliashberg equations [31].
- **no proximity effect:** A thin layer of a noble metal on top of a superconductor can be described by means of a generalization of Eliashberg equations [14].

8 Cuprate high-temperature superconductors

The standard Eliashberg equations have been derived for superconductors where the energy gap showed an s -wave symmetry and the Cooper pairing was mediated by the electron-phonon interaction. This type of interaction allows the application of Migdal's theorem, which states that vertex corrections in the electron-phonon interaction can be neglected to order $\lambda\omega_D/E_F$. On the other hand, it is now widely accepted that the high T_c cuprates have an energy gap with d symmetry [26]. As concerns the microscopic mechanism leading to Cooper pairs, in these materials, there is still no consensus, even though there are indications that antiferromagnetic spin fluctuations can play an important role [28, 29]. In principle, different mechanisms can provide a transition to the superconducting state without any phonon participation. But a scenario where superconductivity occurs through a joint contribution of the phonon and electronic mechanisms is also perfectly realistic. If different mechanisms are considered the Migdal's theorem does not work a priori. In this case the vertex corrections, at least in principle, cannot be neglected and a new type of Eliashberg equations could be necessary. The simplest model of cuprates uses a single-band approximation in a two-dimensional case, thus referring, for example, to the a - b planes of the layered superconductors and neglecting the band dispersion and the gap in the c -direction. For simplicity, one can consider the Fermi surface as circle in the a - b plane and the wave vectors \mathbf{k} and \mathbf{k}' completely determined by the respective azimuthal angles ϕ and ϕ' , since their length is, as usual, taken equal to k_F . The d -wave one-band Eliashberg equations in the imaginary axis representation are [32]

$$\omega_n Z(i\omega_n, \phi) = \omega_n + \frac{T}{2} \sum_m \int_0^{2\pi} d\phi' \Lambda(i\omega_n - i\omega_m, \phi, \phi') N_Z(i\omega_m, \phi') + \Gamma \frac{N_Z(i\omega_n)}{c^2 + N_Z(i\omega_n)^2}$$

$$Z(i\omega_n, \phi) \Delta(i\omega_n, \phi) = \frac{T}{2} \sum_m \int_0^{2\pi} d\phi' [\Lambda(i\omega_n - i\omega_m, \phi, \phi') - \mu^*(\phi, \phi', \omega_c) \vartheta(\omega_c - \omega_m)] N_\Delta(i\omega_m, \phi')$$

where ϑ is the Heaviside function, ω_c a cut-off energy,

$$\Lambda(i\omega_n - i\omega_m, \phi, \phi') = 2 \int_0^{+\infty} d\Omega \alpha^2 F(\Omega, \phi, \phi') / [(\omega_n - \omega_m)^2 + \Omega^2],$$

$$N_\Delta(i\omega_m, \phi) = \frac{\Delta(i\omega_m, \phi)}{\sqrt{\omega_m^2 + \Delta^2(i\omega_m, \phi)}},$$

$$N_Z(i\omega_m, \phi) = \frac{\omega_m}{\sqrt{\omega_m^2 + \Delta^2(i\omega_m, \phi)}},$$

and

$$N_Z(i\omega_n) = \frac{1}{2\pi} \int_0^{2\pi} N_Z(i\omega_n, \phi) d\phi.$$

Γ is proportional to the impurity concentration or disorder and c is related to the electron phase shift for scattering off an impurity. The Born limit is found when $c = \infty$ but Γ/c^2 constant, and the unitary limit when $c = 0$. In the Born limit there is a different behavior between s - and d -symmetries: in the s -wave symmetry a non magnetic impurities do not affect the critical

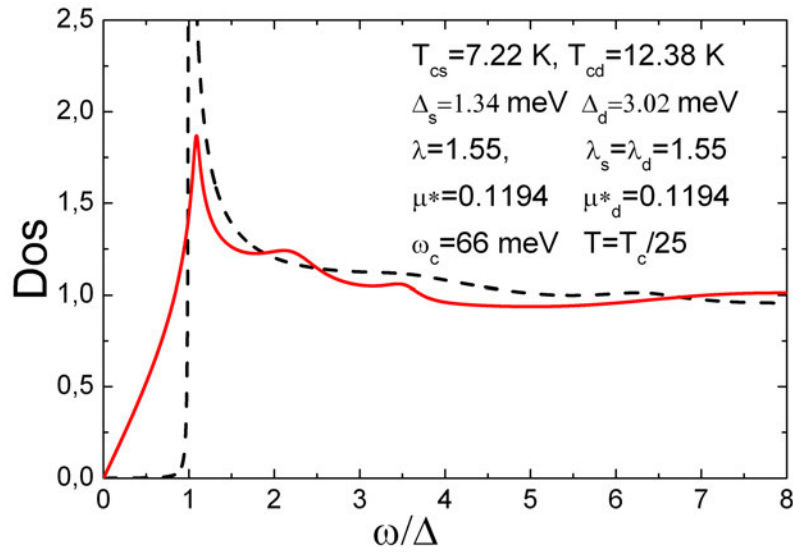


Fig. 4: The calculated superconductive density of states in the d -wave (red solid line) and s -wave (black dashed line) case.

temperature while in the d -wave symmetry the critical temperature is strongly reduced. In the simplest model the electron-boson spectral function $\alpha^2(\Omega)F(\Omega, \phi, \phi')$ and the Coulomb pseudopotential $\mu^*(\phi, \phi')$ to lowest order contain both s - and d -wave contributions,

$$\alpha^2 F(\Omega, \phi, \phi') = \alpha^2 F_s(\Omega) + \alpha^2 F_d(\Omega) \sqrt{2} \cos(2\phi) \sqrt{2} \cos(2\phi') \quad (53)$$

$$\mu^*(\phi, \phi') = \mu_s^* + \mu_d^*(\Omega) \sqrt{2} \cos(2\phi) \sqrt{2} \cos(2\phi'). \quad (54)$$

A solution with a pure d -wave gap function $\Delta(\omega, \phi') = \Delta_d(\omega) \cos(2\phi)$ and a pure s -wave renormalization function $Z(\omega, \phi') = Z_s(\omega)$ can be obtained. Indeed the equation for $Z_d(\omega)$ is a homogeneous integral equation whose only solution in the weak-coupling regime is $Z_d(\omega) = 0$. Even though in the strong-coupling limit a non-zero solution could exist above a certain coupling strength threshold, but usually one does not consider this rather exotic case and then the stable solution corresponds to $Z_d(\omega) = 0$ for all the couplings.

By assuming d -wave symmetry for the gap function, the parameter μ_s^* does not enter into the two relevant Eliashberg equations. Therefore, although it is certainly larger than μ_d^* and so drives the system towards d -wave symmetry, it does not influence the solution.

The superconductive density of states $N_d(\omega)$ can be easily obtained from the calculated frequency-dependent gap function as

$$N_d(\omega) = \int_0^{2\pi} \frac{d\phi}{2\pi} \Re \left\{ \frac{\omega}{\sqrt{\omega^2 - \Delta_d^2(\omega) \cos^2(2\phi)}} \right\}. \quad (55)$$

In Fig. 4 the superconductive density of states calculated in s - and d -wave cases are compared. The black-dotted line is the solution obtained for lead, the red-solid line is an ideal solution where the input parameters are kept the same of the case of lead, but d -wave symmetry has been imposed.

9 Multi-band Eliashberg theory

The equations seen so far (in all their formulations) are suitable to describe only a relatively small number of superconductors. There are many materials which are less trivial and show a multi-band structure. Consider a superconductor containing several different groups of electrons occupying distinct quantum states. The most typical example is a material with several overlapping energy bands. One can expect that each band will possess its own energy gap. This means that the density of states of the superconducting pairs contains several peaks. Of course if the energy gap were defined as the smallest quantum of energy that can be absorbed by the material, then only the smallest gap of the system would satisfy this definition. Thus to avoid misunderstandings when talking about the multi-gap structure of a spectrum we will mean explicitly the aforementioned multi-peak property of the density of states. For this case the previous equations must be generalized.

Considering a two band system [25] as the MgB_2 , the parameters multiply: there are now four separate electron-phonon spectral functions $\alpha_{ij}^2 F(\Omega)$ and four Coulomb pseudopotentials μ_{ij}^* , where $i, j = 1, 2$.

The isotropic Eliashberg equations generalized to n bands ($i = 1, \dots, n$), are written on the imaginary axis as

$$\begin{aligned} \omega_n Z_i(i\omega_n) &= \omega_n + \pi T \sum_{mj} \lambda_{ij}(i\omega_n, i\omega_m) N_j^Z(i\omega_m) + \\ &+ \sum_j [\Gamma_{ij} + \Gamma_{ij}^M] N_j^Z(i\omega_n) \end{aligned} \quad (56)$$

$$\begin{aligned} Z_i(i\omega_n) \Delta_i(i\omega_n) &= \pi T \sum_{mj} [\lambda_{ij}(i\omega_n, i\omega_m) - \mu_{ij}^*(\omega_c)] \times \\ &\times \Theta(\omega_c - |\omega_m|) N_j^\Delta(i\omega_m) + \sum_j [\Gamma_{ij} + \Gamma_{ij}^M] N_j^\Delta(i\omega_n) \end{aligned} \quad (57)$$

where Γ_{ij} and Γ_{ij}^M are the non magnetic and magnetic impurity scattering rates, and, in a manner quite similar to the single band case,

$$\lambda_{ij}(i\omega_m - i\omega_n) \equiv 2 \int_0^\infty d\Omega \frac{\Omega \alpha_{ij}^2 F(\Omega)}{\Omega^2 + (\omega_n - \omega_m)^2} \quad (58)$$

and

$$N_j^\Delta(i\omega_m) = \Delta_j(i\omega_m) \cdot \left[\sqrt{\omega_m^2 + \Delta_j^2(i\omega_m)} \right]^{-1}, \quad N_j^Z(i\omega_m) = \omega_m \cdot \left[\sqrt{\omega_m^2 + \Delta_j^2(i\omega_m)} \right]^{-1}.$$

The diagonal elements describe the intra-band coupling, while the off-diagonal the inter-band one. The values of the inter-band coupling constants are not completely free, but there is a constraint

$$\frac{\lambda_{ij}}{\lambda_{ji}} = \frac{N_i(0)}{N_j(0)}. \quad (59)$$

This means that the ratio of the interband coupling constant λ_{12} and λ_{21} is equal to the ratio of density of states.

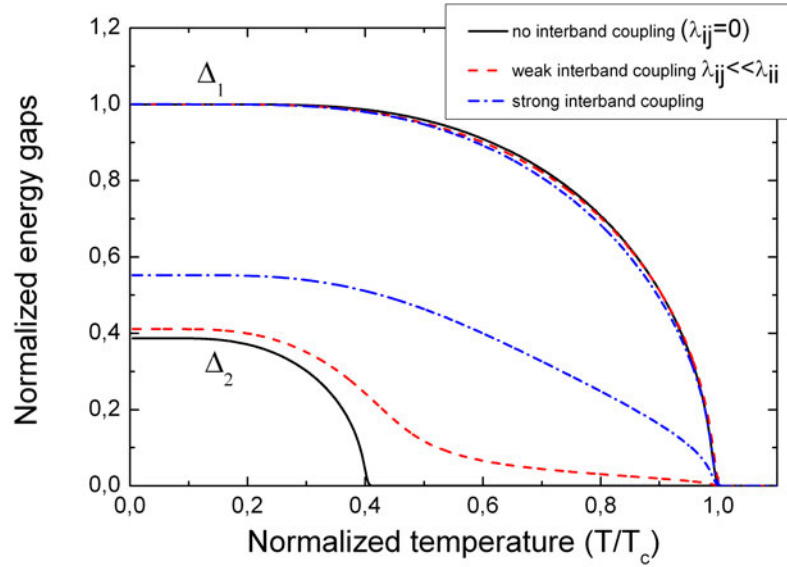


Fig. 5: Temperature dependence of the gaps Δ_1 and Δ_2 in a two-band model, calculated in the cases of (i) no intraband coupling ($\lambda_{ij}=0$), (ii) weak intraband coupling (dotted lines), and (iii) strong interband coupling (dash-dot lines). The intraband coupling constants are arbitrary; here we used those for MgB_2

It is interesting and propedeutic for the subsequent chapters to analyze different situations, as the coupling constants change: the limit of small inter-band coupling and the opposite case, i.e., a pure inter-band case [25] will be considered. The first case is interesting because it allows understanding that an, even small, interband coupling leads to the correlation of the two bands, otherwise completely independent as it is shown Fig. 5.

In a superconductor without interband coupling ($\lambda_{ij} = \lambda_{ji} = 0$) the bands behaves as n different superconductors that have n different transition temperatures, T_{C1} and T_{C2} , each associated with the respective band. The resulting superconducting state that results will be given by the sum of the n bands contributions which are completely independent. As the off-diagonal components grow the n bands become connected. However, this does not means that the superconductor behaves as a one-band system. As long as each band has different a spectral function, and a different coupling constant, they will give different contributions. Changing the off-diagonal elements λ_{ij} results in different temperature of the upper and lower gaps. Each band contains its own set of Cooper pairs. Since, generally speaking, k_{F_i} and k_{F_j} (here k_{F_i} and k_{F_j} are on the Fermi surface for different bands), there is no pairing of electrons belonging to different energy bands i.e. λ_{ij} , of course, does not represent a pairing between electrons of different bands. This does not mean, however, that the pairing within each band is completely insensitive to the presence of the other. On the contrary, a peculiar interband interaction and the appearance of nonlocal coupling constants are fundamental properties of the multiband model. Consider two electrons belonging to band i . They exchange phonons and form a pair as a result. There exist two pairing scenarios. In one of them, the first electron emits a virtual phonon and makes a transition into a state within the same energy band. The second electron absorbs the phonon

and also remains in the same energy band, forming a bound pair with the first one. This is the usual pairing picture, described by a coupling constant λ_{ii} . However, the presence of the other energy band gives rise to an additional channel. Namely, the first electron, originally located in band i , can emit a virtual phonon and make a transition into band k . The phonon is absorbed by the second electron, which also is scattered into band k , where it pairs up with the first electron. As we know, there is no energy conservation requirement for single virtual transitions; such conservation, however, must hold for the initial and final states. In our case this criterion is met. Indeed, the initial and final states correspond to particles on the Fermi surface. Note that, in addition, the initial and final total momenta are equal to zero. Thus the initial state had two electrons in band i , while the final state finds a pair in band k . Interband charge transfer processes are described by nondiagonal coupling constants λ_{ij} , and because of them the system is characterized by a single critical temperature. Otherwise, each band would have its own T_c . There is a formal similarity between the Eliashberg equations for a proximity system [14] and those for a two band system: if the mathematical expression of Eliashberg theory for a system with two gaps is compared with a proximity system it is possible to notice a profound formal analogy between these two situations. In both cases there is induced superconductivity because in the second band, as in a noble metal film, a very weak intrinsic pairing can be chosen so this band alone would not become superconducting. However the mechanisms giving rise to induced superconductivity are very different. In the two band model the systems are “separated” in momentum space and the second band acquires an order parameter thanks to phonon exchange. The phase space for phonons is effectively increased. In the proximity effect, on the other hand, the systems are spatially separated and superconductivity is induced by the tunnelling of Cooper pairs.

The multiband Eliashberg model developed above can also be used to explain the temperature dependence of the upper critical magnetic field [33]. For the sake of completeness, the linearized gap equations in the presence of magnetic field, for a superconductor in the clean limit are reported. In the following, v_{F_j} is the Fermi velocity of the j -th band, and H_{c2} is the upper critical field

$$\begin{aligned}
 \omega_n Z_i(i\omega_n) &= \omega_n + \pi T \sum_{mj} \lambda_{ij}(i\omega_n - i\omega_m) \text{sign}(\omega_m) \\
 Z_i(i\omega_n) \Delta_i(i\omega_n) &= \pi T \sum_{mj} [\lambda_{ij}(i\omega_n - i\omega_m) - \mu_{ij}^*(\omega_c)] \times \\
 &\quad \times \theta(|\omega_c| - \omega_m) \chi_j(i\omega_m) Z_j(i\omega_m) \Delta_j(i\omega_m) \\
 \chi_j(i\omega_m) &= \frac{2}{\sqrt{\beta_j}} \int_0^{+\infty} dq \exp(-q^2) \times \\
 &\quad \times \tan^{-1} \left[\frac{q \sqrt{\beta_j}}{|\omega_m Z_j(i\omega_m)| + i\mu_B H_{c2} \text{sign}(\omega_m)} \right].
 \end{aligned}$$

Here $\beta_j = \pi H_{c2} v_{F_j}^2 / (2\Phi_0)$ and Φ_0 is the unit of magnetic flux. In these equations the bare Fermi velocities v_{F_j} [33] are the input parameters.

10 Iron pnictide superconductors

10.1 Gaps, critical temperature, and upper critical magnetic field

The new class of superconductive Fe-based compounds [34, 35] shows similar characteristics to the cuprates and the heavy fermions, for example the high values of the ratio $2\Delta/T_c$ or the presence of the pseudogap. For all three classes of materials it is proposed that superconductivity is mediated by antiferromagnetic spin fluctuations [28]. The most obvious difference is that almost all the iron compounds present a multiband behavior while in HTCS and in heavy fermions this behavior was detected only in some particular cases. The multi-band nature of Fe-based superconductors may give rise to a multi-gap scenario [36] that is indeed emerging from different experimental data with evidence for rather high gap ratios, $\approx 2 - 3$. In this regard neither a three-band BCS model is adequate (it can only account for the gap ratio and T_c but not for the exact experimental gap values) nor a four-band Eliashberg model with small values of the coupling constants and large boson energies because the calculated critical temperature has turned out to be larger than the experimental one. The high experimental value of the larger gap suggests that high values of the coupling constants might be necessary to explain the experimental data within a three-band model [37]: one has therefore to employ the Eliashberg theory for strong coupling superconductors. At the beginning a three-band Eliashberg model allowed reproducing various experimental data, indicating that these compounds can represent a case of dominant negative interband-channel superconductivity ($s\pm$ wave symmetry) with small typical boson energies (≈ 10 meV) but too high values of the electron-boson coupling constants ($1.9 \leq \lambda_{\text{tot}} \leq 5.9$). The way to solve this problem is suggested by experiments of Inosov and coworkers [38]: they found that the temperature evolution of the spin resonance energy follows the superconducting energy gap and this fact should indicate a *feedback effect* [28, 39, 40] of the condensate on the spin fluctuations.

Then the experimental low temperature spin resonance can be chosen as the representative boson energy and the two remaining free parameters can be fixed in order to reproduce the exact experimental gap values. After this, with the same parameters, the critical temperature T_c^* can be calculated. Generally it is always $T_c^* \gg T_c^{\text{exp}}$ where T_c^{exp} is the experimental critical temperature. In the next step the same temperature dependence of the superconductive gap has to be imposed to the representative boson energy while the other input parameters used before are kept fixed.

Of course, at $T = T_c^*$ the energy peaks of the spectral functions (the representative boson energy) is equal to zero while at $T = 0$ K the new spectral functions are equal to the old ones. In this way, by taking into account the *feedback effect* of the condensate [28, 39, 40] on the antiferromagnetic spin fluctuations it is possible to explain the experimental data (the gap values and the critical temperature) in a model with only two free parameters in a moderate strong coupling regime ($\lambda_{\text{tot}} \approx 1.5 - 2$).

Four representative cases are reported (three hole type and one electron type): $\text{LaFeAsO}_{1-x}\text{F}_x$, $\text{SmFeAsO}_{1-x}\text{F}_x$, $\text{Ba}_{1-x}\text{K}_x\text{Fe}_2\text{As}_2$, and $\text{Ba}(\text{Fe}_x\text{Co}_{1-x})_2\text{As}_2$. The electronic structure of the hole

type compounds can be approximately described by a three-band model with two hole bands (indicated in the following as bands 1 and 2) and one equivalent electron band (3) [37] while for one electron type compound the model contains one hole band (indicated in the following as band 1) and two equivalent electron bands (2 and 3) [41]. The s -wave order parameters of the hole bands Δ_1 and Δ_2 have opposite sign with respect to that of the electron band Δ_3 [42] in the hole type case while Δ_1 has opposite sign with respect to that of the two electron bands, Δ_2 and Δ_3 in the electron type case [41]. In such systems, in the first approximation, intraband coupling could only be provided by phonons ph , and interband coupling only by antiferromagnetic spin fluctuations sf [37, 39]. The experimental data concerning the four compounds considered can be summarized as follow [39]:

1. LaFeAsO_{0.9}F_{0.1} (LaFeAsOF) with $T_c^A = 28.6$ K where point-contact spectroscopy measurements gave $\Delta_1(0) \approx 8.0$ meV and $\Delta_2(0) \approx 2.8$ meV;
2. Ba_{0.6}K_{0.4}Fe₂As₂ (BaKFeAs) with $T_c = 37$ K where ARPES measurements gave $\Delta_1(0) = 12.1 \pm 1.5$ meV, $\Delta_2(0) = 5.2 \pm 1.0$ meV and $\Delta_3(0) = 12.8 \pm 1.4$ meV
3. SmFeAsO_{0.8}F_{0.2} (SmFeAsOF) with $T_c^A = 52$ K ($T_c^{\text{bulk}} = 53$ K) where point-contact spectroscopy measurements gave $\Delta_1(0) = 18 \pm 3$ meV and $\Delta_2(0) = 6.15 \pm 0.45$ meV;
4. Ba(Fe_xCo_{1-x})₂As₂ (BaFeCoAs) with $T_c^A = 22.6$ K ($T_c^{\text{bulk}} = 24.5$ K) where point-contact spectroscopy measurements gave $\Delta_1(0) = 4.1 \pm 0.4$ meV and $\Delta_2(0) = 9.2 \pm 1$ meV.

T_c^A is the critical temperature obtained by Andreev reflection measurements [36] and T_c^{bulk} is the critical temperature obtained by transport measurements. Note that only in the case of ARPES the gaps are associated to the relevant band since point-contact spectroscopy measurements generally gives only two gaps, the larger one has been arbitrarily indicated as Δ_1 supposing that $\Delta_1 \sim |\Delta_3|$. To obtain the gap-values and the critical temperature within the $s \pm$ wave, one has to solve six coupled equations for the gaps $\Delta_i(i\omega_n)$ and the renormalization functions $Z_i(i\omega_n)$, where i is the index of the bands and ranges from 1 to 3. The solution of the system of multiband Eliashberg equations requires a huge number of input parameters (18 functions and 9 constants); however, some of these parameters are related to one another, some can be extracted from experiments and some can be fixed by suitable approximations. In the case of the pnictides several assumptions can be made: (i) the total electron-phonon coupling constant is small, (ii) phonons mainly provide intraband coupling [42], (iii) spin fluctuations mainly provide interband coupling [42]. The simplest way to take these assumptions into account is to set $\lambda_{ii}^{ph} = \lambda_{ij}^{ph} = \lambda_{ii}^{sf} = 0$. Indeed the upper limit of the phonon coupling in these compounds is ≈ 0.35 [43], and the intraband spin-fluctuation coupling can be neglected [42]. Moreover, the phonon couplings and the Coulomb pseudopotentials roughly compensate each other, then $\mu_{ii}^*(\omega_c) = \mu_{ij}^*(\omega_c) = 0$ [37]. Within these approximations, the electron-boson coupling-constant matrix λ_{ij} becomes [37, 39, 41]:

$$\lambda_{ij} = \begin{pmatrix} 0 & \lambda_{12} & \lambda_{13} \\ \lambda_{21} = \lambda_{12}\nu_{12} & 0 & \lambda_{23} \\ \lambda_{31} = \lambda_{13}\nu_{13} & \lambda_{32} = \lambda_{23}\nu_{23} & 0 \end{pmatrix} \quad (60)$$

	λ_{tot}	$\lambda_{\text{tot}}^{\text{old}}$	$\lambda_{12/21}$	$\lambda_{13/31}$	$\lambda_{23/32}$	Ω_0 (meV)
BaFeCoAs	1.87		0.76/0.85	1.21/5.44	0.00/0.00	9.04
	2.83	1.93	0.91/1.02	2.08/9.35	0.00/0.00	9.04
	1.72		0.77/0.87	1.05/4.72	0.00/0.00	9.04
	2.52		0.90/1.01	1.76/7.91	0.00/0.00	9.04
	1.75		0.00/0.00	2.11/1.91	0.40/0.21	11.44
LaFeAsOF	2.38	2.53	0.00/0.00	2.93/2.66	0.46/0.24	11.44
	2.04		0.00/0.00	2.27/2.27	0.56/0.28	14.80
BaKFeAs	2.84	3.87	0.00/0.00	3.21/3.21	0.67/0.34	14.80
	1.72		0.00/0.00	1.55/3.88	0.42/0.84	20.80
SmFeAsOF	2.39	5.90	0.00/0.00	2.23/5.58	0.49/0.98	20.80

Table 2: The values of Ω_0 and λ_{ij} , that allow reproducing the experimental gap values, are shown. λ_{tot} is compared with $\lambda_{\text{tot}}^{\text{old}}$ that is the value determined in the previous works [37, 41]. In the first rows the *sf* spectral functions used have usual shape while in the second ones have Lorentzian shape.

where $\nu_{ij} = N_i(0)/N_j(0)$ and $N_i(0)$ is the normal density of states at the Fermi level for the i -th band. In the hole case $\lambda_{21} = \lambda_{12} = 0$ while in the electron case $\lambda_{23} = \lambda_{32} = 0$.

In the numerical simulations the standard form for the antiferromagnetic spin fluctuations is used: $\alpha_{ij}^2 F^{sf}(\Omega) = B_{ij} \Omega \Omega_{ij} \Theta(\Omega_{\text{max}} - \Omega) / (\Omega^2 + \Omega_{ij}^2)$ where B_{ij} are the normalization constants necessary to obtain the proper values of λ_{ij} while Ω_{ij} are the peak energies. In all the calculations, for simplicity, $\Omega_{ij} = \Omega_0$. The maximum spin-fluctuation energy is $\Omega_{\text{max}} = 10\Omega_0$, the cut-off energy is $\omega_c = 30\Omega_0$ and the maximum quasiparticle energy is $\omega_{\text{max}} = 40\Omega_0$. The typical *sf* energy Ω_0 is the spin resonance energy that has been measured and the empirical relation $\Omega_0 = (2/5)T_c$ available in literature [44] is assumed to be correct for all compounds examined.

Bandstructure calculations provide information about the factors ν_{ij} that enter the definition of λ_{ij} . In the case of $\text{LaFeAsO}_{0.9}\text{F}_{0.1}$, $\nu_{13} = 0.91$ and $\nu_{23} = 0.53$; for $\text{Ba}_{0.6}\text{K}_{0.4}\text{Fe}_2\text{As}_2$, $\nu_{13} = 1$ and $\nu_{23} = 2$; in $\text{SmFeAsO}_{0.8}\text{F}_{0.2}$, $\nu_{13} = 0.4$ and $\nu_{23} = 0.5$ and in $\text{Ba}(\text{Fe}_x\text{Co}_{1-x})_2\text{As}_2$, $\nu_{12} = 1.12$ and $\nu_{13} = 4.50$ [39].

First of all the imaginary-axis Eliashberg equations are solved in order to calculate the low-temperature values of the gaps (which are actually obtained by analytical continuation to the real axis by using the technique of the Padé approximants) and so the two free parameters

	$\Delta_1(\text{meV})$	$\Delta_2(\text{meV})$	$\Delta_3(\text{meV})$	$T_c(\text{K})$	$T_c^*(\text{K})$
BaFeCoAs	6.63	-4.07	-9.18	26.07	33.00
	7.02	-4.12	-9.18	23.73	28.95
	5.89	-3.78	-8.17	23.43	29.69
	6.19	-3.79	-8.19	21.72	26.41
LaFeAsOF	8.01	2.82	-7.75	29.37	37.22
	8.01	2.77	-7.71	26.86	31.81
BaKFeAs	12.04	5.20	-12.00	43.66	55.26
	12.04	5.24	-11.91	38.33	46.18
SmFeAsOF	14.86	6.15	-18.11	58.53	74.13
	15.51	6.15	-18.00	52.80	63.82

Table 3: *The calculated values of the gaps and of the two critical temperature with and without feedback effect. In the first rows the sf (spin-fluctuation) spectral functions used have the usual shape while in the second line they have Lorentzian shape.*

of the model are fixed: λ_{13} and λ_{23} (λ_{12}). By properly selecting the values of λ_{13} and λ_{23} (λ_{12}) it is relatively easy to obtain the experimental values of the gaps with reasonable values of $\lambda_{tot} = \sum_{ij} N_i(0)\lambda_{ij} / \sum_{ij} N_i(0)$ (between 1.72 and 2.04). However, in all the materials examined, the high $2\Delta_{1,3}/k_B T_c$ ratio (of the order of 8-9) makes it possible to reproduce *also* the values of the large gap(s) only if the calculated critical temperature T_c^* is considerably higher than the experimental one. For solving this problem, that is also present in the HTCS, it is necessary to assume the existence a feedback effect [28, 39] of the condensate and, in a phenomenological way, a temperature dependence of the representative boson energy $\Omega_0(T) = \Omega_0 \tanh(1.76\sqrt{T_c^*/T - 1})$ that is, approximately, the temperature dependence of the gap in the strong coupling case, is introduced in the Eliashberg equation.

The primary effect of this assumption is the reduction of the critical temperature without changes in the gap values at $T \ll T_c^*$. For a completely consistent procedure it should be used $\Omega_0(T) = \Omega_0 \eta(T)$ where $\eta(T)$ is the temperature dependent part of the superfluid density $\rho(T) = \rho(0)\eta(T)$ and $\rho(0)$ is the superfluid density at $T = 0$ K. $\eta(T)$ is a function of $\Delta_i(i\omega_n)$ and so, in this way, the numerical solution of the Eliashberg equations becomes remarkably more complex and time consuming.

For a general picture of the physical landscape the Eliashberg equations have to be solved in three different situations: (i) only *sf* inter-band coupling is present and the *sf* spectral functions

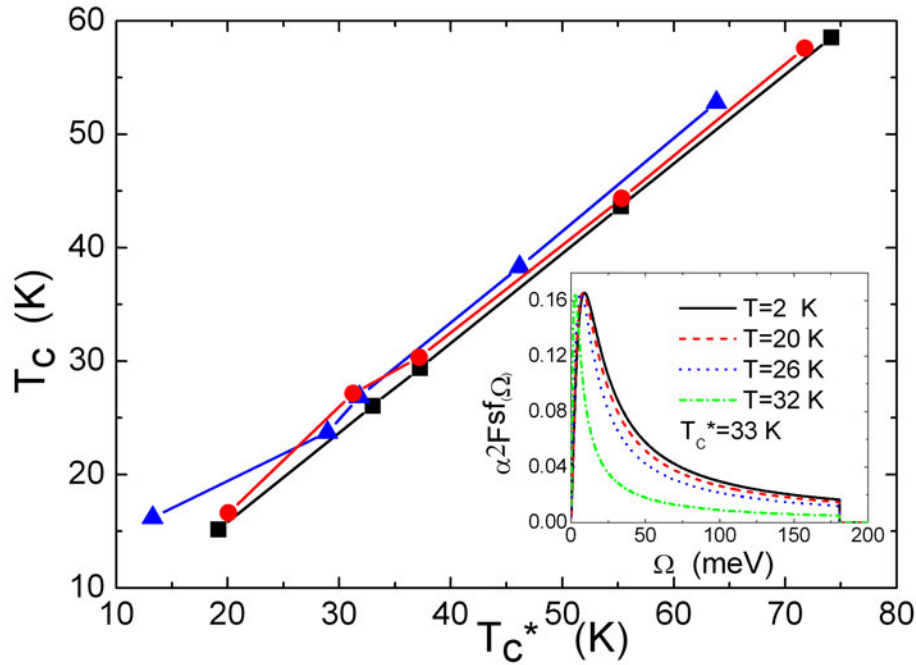


Fig. 6: Calculated critical temperature T_c with feedback effect versus standard critical temperature T_c^* in three different situations: only interband sf coupling with standard spectral functions (black squares), interband sf coupling with standard spectral functions and small intraband ph coupling (red circles) and only interband sf coupling with Lorentz spectral functions (dark blue triangles). The insert shows the sf spectral function, for $Ba(Fe_xCo_{1-x})_2As_2$ at different temperatures ($T < T_c^*$) with the feedback effect.

have the usual shape of the normal state (ii) sf interband coupling with a small ph intra-band contribution is present and sf spectral functions have the usual shape, and (iii) only sf interband coupling is present and the sf spectral functions have Lorentz shape.

In the first case the coupling constant λ_{tot} is in the range 1.72-2.04. The results are almost independent from Ω_{max} . The agreement with the experimental critical temperature is good. It is important to notice that the coupling parameters almost do not change in these considered case. In the second case there is also an intra-band phonon contribution, equal in each band and in each compound for simplicity, with $\lambda_{ii}^{ph} = 0.35$ and $\Omega_0^{ph} = 18$ meV that are the upper limits for the ph coupling constants and the representative ph energies [43]. The ph spectral functions have Lorentzian shape [37]. The phonon peaks are all in $\Omega_{ij} = \Omega_0^{ph}$, and the antiferromagnetic spin fluctuations peak in $\Omega_{ij} = \Omega_0^{sf}$ and the half-width⁴ is always 2 meV [39] and $\omega_c = 12\Omega_0^{ph}$. λ_{tot} and T_c are practically the same as the previous case. This last fact indicates that the effect of intraband phonon contributions is negligible. In the third case (Lorentz shape of sf spectral functions) the agreement with the experimental critical temperatures is very good in all compounds but the total coupling is larger ($2.38 \leq \lambda_{\text{tot}} \leq 2.84$).

⁴In more recent work the half-width of the Lorentzian functions describing antiferromagnetic spin fluctuations is equal to $\Omega_0^{sf}/2$ [38].

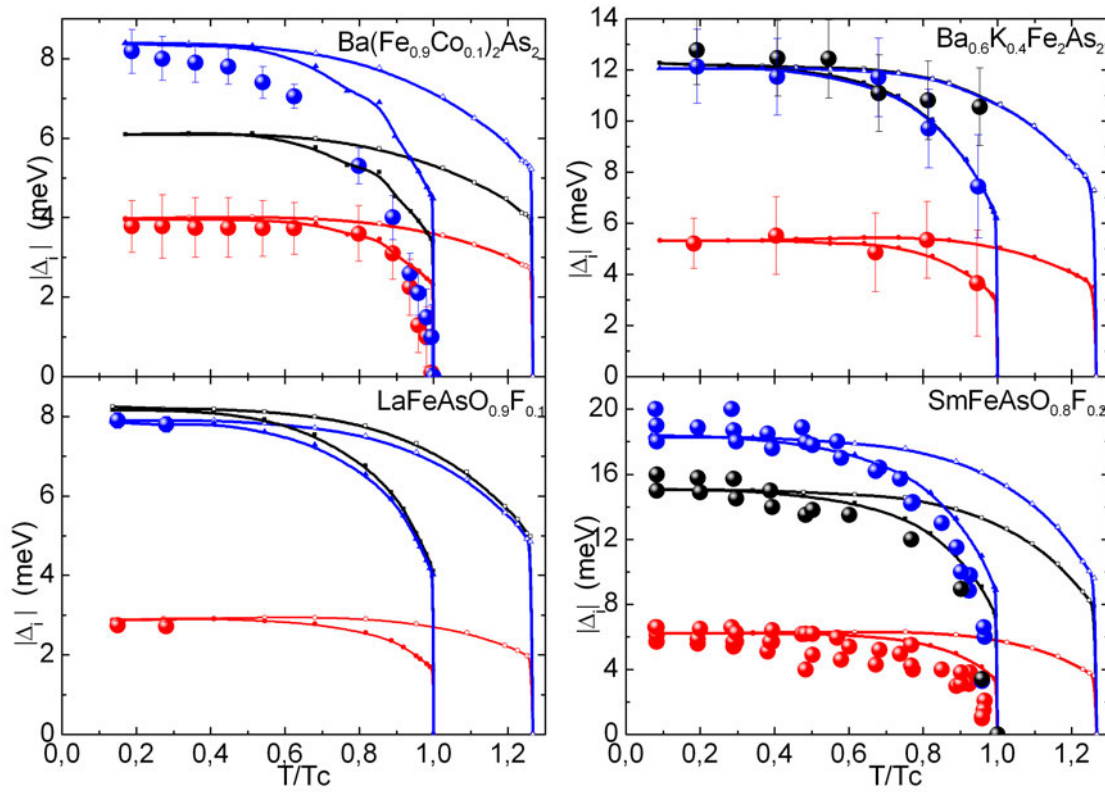


Fig. 7: Calculated temperature dependence of $|\Delta_i|$ from the solution of real axis Eliashberg equations in the standard case (open symbol) and when the feedback effects are present (solid symbol): $|\Delta_1|$ black, $|\Delta_2|$ red, and $|\Delta_3|$ dark blue. The experimental data [39] are shown as big solid bullets.

Fig. 6 shows the linear relation between T_c and T_c^* in all three cases examined. In Table 2 the input parameters of the Eliashberg equations in the first and third case examined for the four compounds are listed. Table 3 summarizes the calculated values of the gaps and critical temperatures T_c and T_c^* obtained by numerical solution of the Eliashberg equations. Once the values of the low-temperature gaps were obtained, the temperature dependence can be calculated by directly solving the three-band Eliashberg equations in the real-axis formulation instead of using the analytical continuation to the real axis of the imaginary-axis solution. Of course, the results of the two procedures are virtually identical at low temperature.

In Fig. 7 the calculated temperature dependence of $|\Delta_i|$ is compared with the experimental data and the agreement is very good. In all cases, their behavior is rather unusual and completely different from the BCS since the gaps slightly decrease with increasing temperature until they suddenly drop close to T_c . This arises from a complex non-linear dependence of the Δ vs. T curves on the λ_{ij} values and is possible only in a strong-coupling regime [45]. Curiously, in all four compounds the rate T_c^*/T_c is the same, 1.27. The three-band Eliashberg equations for the upper critical field [40] are solved. Here the three bare Fermi velocities v_{Fj} are the input parameters. The number of adjustable parameters can be reduced to one by assuming that, as in

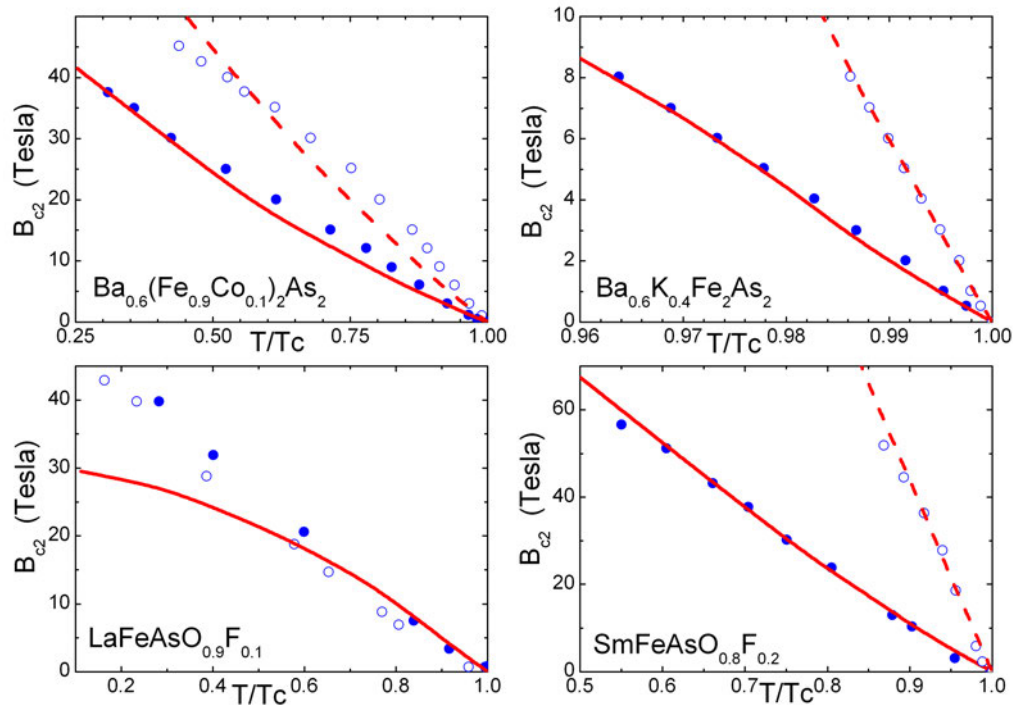


Fig. 8: Experimental temperature dependence of the upper critical field in $\text{Ba}(\text{Fe}_x\text{Co}_{1-x})_2\text{As}_2$, $\text{LaFeAsO}_{0.9}\text{F}_{0.1}$, $\text{Ba}_{0.6}\text{K}_{0.4}\text{Fe}_2\text{As}_2$, and $\text{SmFeAsO}_{0.8}\text{F}_{0.2}$ (symbols), and the relevant fitting curves (lines) obtained by solving the Eliashberg equations. $\mathbf{H} \parallel ab$: Solid symbols, solid line and $\mathbf{H} \parallel c$: open symbols, dashed line.

a free-electron gas, $v_{F_j} \propto N_j(0)$ so that $v_{F_2} = v_{F_1}\nu_2/\nu_1$ and $v_{F_3} = v_{F_1}/\nu_1$, thus leaving v_{F_1} as the only free parameter.

Fig. 8 depicts the experimental values of the upper critical field measured [40] in the case of $\text{Ba}(\text{Fe}_x\text{Co}_{1-x})_2\text{As}_2$, $\text{Ba}_{0.6}\text{K}_{0.4}\text{Fe}_2\text{As}_2$, $\text{LaFeAsO}_{0.9}\text{F}_{0.1}$, and $\text{SmFeAsO}_{0.8}\text{F}_{0.2}$ compared to the best-fitting curve obtained by solving the Eliashberg equations as discussed above. The quality of the fit is rather good in almost all cases, which is a remarkable result of the model in spite of the crudeness of the free-electron approximation. The phenomenology of iron-pnictides superconductors can be explained in the framework of a three-band $s\pm$ wave Eliashberg theory with only two free parameters plus a *feedback effect* i.e., the effect of the condensate on the antiferromagnetic spin fluctuations responsible of the superconductivity in these compounds. Indeed in the four iron compounds discussed, it is possible to reproduce the experimental critical temperature, the gap values and the upper critical field in a moderate strong-coupling regime: $\lambda_{\text{tot}} \approx 1.7 - 2.0$. The large value of the ratio between the gaps and the critical temperature finds a natural justification in this model. Eventually, for describing the phenomenology of the iron compound LiFeAs , it is necessary to use $s\pm$ four-band Eliashberg equations in a moderate strong coupling regime $\lambda_{\text{tot}} = 1.6$ and also in this case the agreement with the experimental data is good [46].

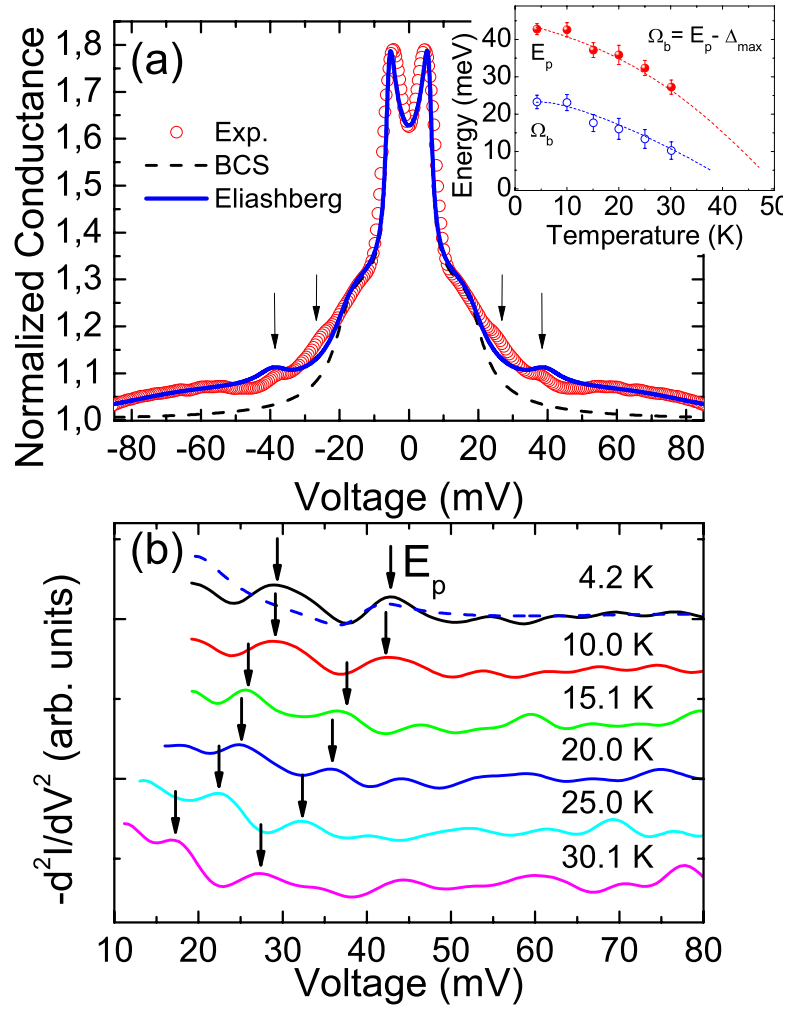


Fig. 9: (a) Normalized experimental conductance curve (circles) obtained in a $\text{Ag/SmFeAsO}_{0.8}\text{F}_{0.2}$ point-contact. The dashed line is a BTK fit to the experiment, obtained using the constant BCS values for the gaps. The solid line is a theoretical curve obtained by introducing in the BTK model the energy-dependent gap functions calculated within the three-band Eliashberg theory. (b) Temperature dependence of the $-d^2I/dV^2$ curves obtained from the same contact as in (a), showing the displacement of the bosonic structures with increasing temperature. The dashed line is obtained from the theoretical curve shown in (a). Inset: temperature dependence of the energy peak, E_p (full symbols) extracted from (b) together with the corresponding boson energy $\Omega_b(T) = E_p(T) - \Delta_{\max}(T)$. Lines are guides to the eye.

10.2 Interaction mechanism

In moderate- or strong-coupling superconductors the tunneling or the Andreev reflection [36] conductance curves can show signatures of the energy dependence of the superconducting gap. These structures are more easily observable if the amplitude of the Andreev signal is large; their signature in the second derivative of the I - V curve can be related to the electron-boson spectral function. In the following, results obtained in an iron compound that provide examples of such strong-coupling effects and of their analysis will be shown, they strongly support a

spin-fluctuation-mediated origin of superconductivity if compared with the inelastic neutron scattering measurements.

Fig. 9 shows a normalized experimental conductance curve obtained on an optimally doped $\text{SmFeAsO}_{0.8}\text{F}_{0.2}$ polycrystal (circles). The amplitude of the Andreev signal in this contact is exceptionally high [36] (about 80%) and, in addition to the clear two-gap features (peaks and shoulders), additional structures or small kinks can be seen around 27 and 40 meV. The dashed line is a BTK [36] fit to the experiment using the two-band 2D BTK model with BCS gap values (i.e. independent of energy). The fit reproduces very well the experiment in the central part of the curve (and allows obtaining reliable values of the gaps) but fails at higher energies. The solid line is instead the result of inserting in the same BTK model the energy-dependent order parameters obtained by solving the three-band Eliashberg equations in which, as usual, the electron-boson spectral function is modeled by a Lorentzian curve [36]. Since to the best of the present knowledge no spin-resonance energy value is available for this compound, the characteristic energy has been chosen, following [44], by extrapolating the relationship $\Omega_0 = 2T_c/5 \approx 20$ meV. Although the theoretical curve shows no structure at 27 meV, the feature at 40 meV is remarkably well reproduced as can be observed in Fig. 9 (a) and (b) (dashed line). Only the structure present at approximately $\Omega_0 + \Delta_{max}$ is reproduced and this indicates that the model has to be investigated further or that additional features of the spectral function are playing an important role. As expected, both structures shift in energy on increasing temperature, partly because the amplitude of the superconducting gaps is also decreasing.

The inset to Fig. 9 (a) reports the position of the energy peak in the second derivative, E_p (full symbols) and the values of the characteristic energy of the boson spectrum $\Omega_b = E_p - \Delta_{max}$ (open symbols) as a function that decreases in temperature. This means that Ω_b cannot be the energy of a phonon mode (in that case it would not tend to zero!) and thus rules out a phononic origin of this feature. Instead, the trend of Ω_b is very similar to that of the spin-resonance energy peak reported in the paper of Inosov *et al.* [38] and thus strongly supports a spin-fluctuation mediated pairing mechanism in these compounds.

This three bands Eliashberg model has only two free parameter λ_{13} and λ_{23} and it is able to explain the values of the gaps, the structures after the gaps in the point-contact spectra the critical temperature, the dependence of temperature of the upper critical field etc. Similar results can be obtained [41] for $\text{Ba}(\text{Fe}_x\text{Co}_{1-x})_2\text{As}_2$.

11 Conclusion

The theory of Eliashberg is, in principle, a theory without free parameters, because the two input parameters, the electron-phonon spectral function and the Coulomb pseudopotential can be calculated via the Density Functional Theory (DFT). From the solution of the Eliashberg equations on the real axis all the physical observables can then be calculated through, almost always, simple functions of complex quantities $\Delta(\omega, T)$ and $Z(\omega, T)$ that, of course, have to be calculated numerically. Obviously, in the majority of cases, the materials of interest can not be described by a *s*-wave model. However, using the appropriate generalizations, the phenomenol-

ogy of almost all superconductors (except HTCS underdoped for now) can be reproduced by this formalism. This theory, strictly speaking, is neither a theory from first principles nor a purely phenomenological theory, and this is both its strength – it has a very close relationship with the experimental observation – and its weakness – no one has ever discovered a new superconductor based on the Eliashberg theory.⁵ To be sure, no one has ever discovered a new superconductor reasoning on any theory.

Acknowledgment

I thank Sara Galasso for valuable help.

⁵The main reason is that the energies involved are too small and the accuracy required cannot be achieved in the calculations.

References

- [1] J. Bardeen, L.N. Cooper, J.R. Schrieffer, *Phys. Rev. B* **108**, 1175 (1957)
- [2] L. Pitaevskii: *Phenomenology and Microscopic Theory: Theoretical Foundations* in K.H. Bennemann and John B. Ketterson (eds.): *Superconductivity* (Springer, 2008)
- [3] W. Jones and N.H. March, *Theoretical Solid State Physics* (Dover Publications, 1973)
- [4] A.B. Migdal, *Sov. Phys. JETP* **34**, 996 (1958)
- [5] G.M. Eliashberg, *Sov. Phys. JETP* **11**, 696 (1960); *Sov. Phys. JETP* **12**, 1000 (1961)
- [6] Y. Nambu, *Phys. Rev.* **117**, 648 (1960)
- [7] V. Ambegaokar: *The Green's functions method* in R.D. Parks (ed.): *Superconductivity*, Vol. I (Dekker, New York, 1969)
- [8] P.B. Allen and B. Mitrović: *Theory of Superconducting T_c* in *Solid State Physics*, Vol. **37**, (1983)
- [9] D.J. Scalapino: *The Electron-phonon interaction and strong-coupling superconductor* in R.D. Parks (ed.): *Superconductivity*, Vol. I (Dekker, New York, 1969)
- [10] J.P. Carbotte, *Rev. Mod. Phys. B* **62**, 1027 (1990)
- [11] J.P. Carbotte and F. Marsiglio: *Electron-Phonon Superconductivity* in K.H. Bennemann and John B. Ketterson (eds.): *Superconductivity* (Springer, 2008)
- [12] E.R. Margine and F. Giustino, *Phys. Rev. B* **87**, 024505 (2013)
- [13] F. Marsiglio, *J. Low Temp. Phys.* **87**, 659 (1992)
- [14] V.Z. Kresin, H. Morawitz, S.A. Wolf: *Mechanisms of Conventional and High T_c Superconductivity* (Oxford University Press, 1993)
- [15] P.B. Allen and R.C. Dynes, *Phys. Rev. B* **12**, 905 (1975)
- [16] H.J. Vidberg and J.W. Serene, *J. Low Temp. Phys.* **29**, 179 (1977);
C.R. Leavens and D.S. Ritchie, *Solid State Communication*, **53**, 137, (1985)
- [17] D.S. Sholl, J.A. Steckel: *Density Functional Theory* (Wiley, 2009);
E. Engel, R.M. Dreizler: *Density Functional Theory* (Springer, Heidelberg, 2011)
- [18] I. Giaever and K. Megerle, *Phys. Rev. Lett.* **122**, 1101 (1961)
- [19] E.L. Wolf, *Principles of electron tunneling spectroscopy* (Oxford University Press, 1985)

- [20] W.L. Mc Millan and J.M. Rowell, Phys. Rev. Lett. **14**, 108 (1965);
Phys. Rev. Lett. **178**, 897 (1968)
- [21] W.L. Mc Millan and J.M. Rowell, *Tunneling and strong-coupling superconductivity*
in R.D. Parks (ed.): *Superconductivity*, Vol. I (Dekker, New York, 1969)
- [22] A.A. Galkin, A.I. D'yachenko, V.M. Svistunov, Sov. Phys.-JETP **39**, 1115 (1974)
- [23] F. Marsiglio, T. Startseva and J.P. Carbotte, Phys. Lett. A **245**, 172 (1998)
- [24] L. Pietronero, S. Strässler and C. Grimaldi, Phys. Rev. B **52**, 10516 (1995);
C. Grimaldi, L. Pietronero and S. Strässler, Phys. Rev. B **52**, 10530 (1995)
- [25] E.J. Nicol and J.P. Carbotte, Phys. Rev. B **71**, 054501 (2005)
- [26] E. Schachinger and J.P. Carbotte:
Extended Eliashberg theory for d-wave superconductivity and application to cuprate
in J.K. Sadasiva and S.M. Rao (eds.): *Models and Methods of High-Tc Superconductivity*,
Vol. II (Nova Science Publisher, 2003)
- [27] M.L. Kubic, Physics Reports **338**, 1 (2000); E.G. Maksimov, M.L. Kubic, and O.V. Dolgov,
Advances in Condensed Matter Physics **2010**, 423725 (2010)
- [28] A.V. Chubukov, D. Pines, and J. Schmalian:
A Spin Fluctuation Model for d-Wave Superconductivity
in K.H. Bennemann and John B. Ketterson (eds.): *Superconductivity* (Springer, 2008)
- [29] D. Manske, I. Eremin, and K.H. Bennemann:
Electronic Theory for Superconductivity in High-Tc Cuprates and Sr₂RuO₄
in K.H. Bennemann and John B. Ketterson (eds.): *Superconductivity* (Springer, 2008)
- [30] F. Jutier, G.A. Ummarino, J.C. Griveau, F. Wastin, E. Colineau, J. Rebizant, N. Magnani
and R. Caciuffo, Phys. Rev. B **77**, 024521 (2008)
- [31] A.A. Golubov, I.I. Mazin, Phys. Rev. B **55**, 15146 (1997)
- [32] G.A. Ummarino, R. Caciuffo, H. Chudo, S. Kambe Phys. Rev. B **82**, 104510 (2010);
D. Daghero, M. Tortello, G.A. Ummarino, J.-C. Griveau, E. Colineau, A.B. Shick, R.
Caciuffo, J. Kolorenc and A.I. Lichtenstein, Nature Communications **3**, 786, (2012)
- [33] S.V. Shulga, S.L. Drechsler, G. Fuchs, and K.H. Müller, K. Winzer, M. Heinecke, and
K. Krug, Phys. Rev. Lett. **80**, 1730-1733 (1998); H. Suderow, V.G. Tissen, J.P. Brison,
J.L. Martinez, S. Vieira, P. Lejay, S. Lee and S. Tajima, Phys. Rev. B **70** 134518 (2004)
- [34] Y. Kamihara, T. Watanabe, M. Hirano and H. Hosono, J. Am. Chem. Soc. **130** 3296 (2008)
- [35] G.R. Stewart, Rev. Mod. Phys. **83**, 1589 (2011)

- [36] D. Daghero, M. Tortello, G.A. Ummarino and R.S. Gonnelli, Rep. Prog. Phys. **74**, 124509 (2011)
- [37] G.A. Ummarino, M. Tortello, D. Daghero and R.S. Gonnelli, Phys. Rev. B **80**, 172503 (2009)
- [38] D.S. Inosov, J.T. Park, P. Bourges, D.L. Sun, Y. Sidis, A. Schneidewind, K. Hradil, D. Haug, C.T. Lin, B. Keimer and V. Hinkov, Nature Phys. **6**, 178-181 (2010)
- [39] G.A. Ummarino, Phys. Rev. B **83**, 092508 (2011)
- [40] G.A. Ummarino, J. Supercond. Nov. Magn. **25**, 1333 (2012)
- [41] M. Tortello, D. Daghero, G.A. Ummarino, A.V. Stepanov, J. Jiang, J.D. Weiss, E.E. Hellstrom, and R.S. Gonnelli, Phys. Rev. Lett **105**, 237002 (2010)
- [42] P.J. Hirschfeld, M.M. Korshunov and I.I. Mazin, Rep. Prog. Phys. **74**, 124508 (2011)
- [43] L. Boeri, M. Calandra, I.I. Mazin, O.V. Dolgov and F. Mauri, Phys. Rev. B **82**, 020506R (2010)
- [44] J. Paglione and R.L. Greene, Nature Phys. **6**, 645 (2010)
- [45] G.A. Ummarino and R.S. Gonnelli, Physica C **328**, 189 (1999)
- [46] G.A. Ummarino, S. Galasso and A. Sanna, J. Phys.: Condens. Matter **25**, 205701 (2013)

14 Path Integral Methods for Continuum Quantum Systems

David M. Ceperley

Physics Department

University of Illinois at Urbana-Champaign

1110 W. Green St., Urbana IL, 61801, USA

Contents

1	The path integral formalism	2
2	Quantum statistics with path integrals	5
2.1	Bosons	6
2.2	Fermions	8
2.3	Restricted path integral method	8
3	Exchange of localized particles	10
4	PIMC calculations of supersolid helium	13
5	Lexicon of the quantum-classical isomorphism	13

This lecture gives a brief overview of the path integral picture of degenerate quantum systems. The path integral method is explicitly formulated at non-zero temperature. Including effects of temperature in calculations is important because many, if not most, measurements and practical applications involve significant thermal effects. One might think that to do calculations at a non-zero temperature, we would have to explicitly sum over excited states. Such a summation would be difficult to accomplish once the temperature is above the energy gap because there are so many possible excitations in a many-body system. As we will see, path integral methods do not require an explicit sum over excitations. As an added bonus, they provide an interesting and enlightening window through which to view quantum systems. For fermion systems we encounter, however, the sign problem. The fixed-node approximation can be used to solve it. We first start by introducing imaginary time path integrals for distinguishable particles, i.e., particles without Bose or Fermi statistics. We then discuss the generalization to Bose and Fermi statistics, and consider how this applies to superfluid bosonic systems and to exchange in quantum crystals (solid ^3He and super-solid ^4He). I am only going to discuss the continuum models; many other authors have discussed the equivalent methods for lattice models. Much of the material comes from a chapter in the book *Interacting Electrons* to be published by Cambridge University Press [1].

1 The path integral formalism

To introduce path integrals, we first review properties of the thermal N -body density matrix. The coordinate space representation is defined in terms of the exact N -body eigenstates $\Phi_i(R)$ and energies E_i

$$\rho(R, R'; \beta) = \sum_i \Phi_i^*(R) e^{-\beta E_i} \Phi_i(R'). \quad (1)$$

Here $R = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ is the $3N$ dimensional vector of particle coordinates. In addition to the inverse temperature $\beta = 1/(k_B T)$, the N -body density matrix depends on two sets of N -body coordinates, R and R' . It is “off-diagonal” if $R \neq R'$. The partition function is its trace, the integral over the diagonal density matrix¹

$$Z(\beta) = \int dR \rho(R, R; \beta) = \sum_i e^{-\beta E_i}. \quad (2)$$

Thermodynamic properties are obtained as

$$\langle \mathcal{O} \rangle = \frac{1}{Z(\beta)} \int dR dR' \langle R | \mathcal{O} | R' \rangle \rho(R', R; \beta) \quad (3)$$

or by differentiating the partition function.

The operator identity $\exp(-\beta \mathcal{H}) = [\exp(-\Delta\tau \mathcal{H})]^M$ where $\Delta\tau = \beta/M$, relates the density matrix at a temperature $k_B/\Delta\tau$ to the density matrix at a temperature M times *lower*. Writing

¹This can include tracing over spin or particle number depending on the ensemble.

this identity in the coordinate representation gives

$$\rho(R_0, R_M, \beta) = \int dR_1 \dots dR_{M-1} \prod_{t=1}^M \rho(R_{t-1}, R_t; \Delta\tau). \quad (4)$$

The sequence of intermediate points $\{R_1, R_2, \dots, R_{M-1}\}$ is the *path*, and $\Delta\tau$ is the *time step*. Trotter's formula [2] is a rigorous mathematical result that underpins quantum Monte Carlo: Consider two operators, \hat{T} and \hat{V} . Under general conditions² Trotter's formula holds

$$e^{-\beta(\hat{T}+\hat{V})} = \lim_{n \rightarrow \infty} \left(e^{-\beta\hat{T}/n} e^{-\beta\hat{V}/n} \right)^n. \quad (5)$$

An intuitive justification of this formula is to note that the corrections to an individual term are proportional to the commutator $[\hat{T}/n, \hat{V}/n]$ which scales as $\mathcal{O}(1/n^2)$. The error of the right hand side of Eq. (5) will contain n such corrections so the total error is $\mathcal{O}(1/n)$ and vanishes as $n \rightarrow \infty$. Now take \hat{V} and \hat{T} to be the potential and kinetic³ operators and evaluate them in coordinate space

$$\langle R | e^{-\Delta\tau\hat{V}} | R' \rangle = \exp(-\Delta\tau V(R)) \delta(R - R') \quad (6)$$

$$\langle R | e^{-\Delta\tau\hat{T}} | R' \rangle = (4\lambda\pi\Delta\tau)^{-3N/2} \exp\left(-(R - R')^2/(4\lambda\Delta\tau)\right). \quad (7)$$

Note that we have set $\Delta\tau = \beta/n$ and $\lambda \equiv \hbar^2/2m$. Putting Eq. (6) and (7) together and integrating over the intermediate coordinate⁴ we obtain the so-called “primitive approximation” to the action

$$S_P(R, R'; \Delta\tau) = -\ln \rho(R, R'; \Delta\tau) \approx \frac{3N}{2} \ln(4\pi\lambda\Delta\tau) + \frac{(R - R')^2}{4\lambda\Delta\tau} + \Delta\tau V(R'). \quad (8)$$

Substituting the action, Eq. (8), into the path integral expression, Eq. (4), the partition function is given by

$$Z_D(\beta) = \lim_{M \rightarrow \infty} \int dR_1 \dots dR_M \exp \left[- \sum_{t=1}^M S_P(R_{t-1}, R_t; \beta/M) \right] \quad (9)$$

with the condition $R_0 = R_M$ to obtain the trace. In this formula, Boltzmann or distinguishable particle statistics are assumed and its partition function is written as Z_D . We will consider Bose and Fermi statistics in the next section.

If the potential energy is real, the integrand of Eq. (9) is non-negative and can thus be interpreted as a classical system with an effective classical potential given by the sum in its exponent. This defines an exact mapping of a quantum system onto a classical equilibrium system: the quantum system of N particles in M time slices becomes an NM -particle classical system. The classical system is composed of N “polymers” each having M “beads” with harmonic springs

²In particular if \hat{T} is the non-relativistic kinetic operator and V the Coulomb interaction.

³The kinetic Green's function has to be modified in periodic boundary conditions to make it periodic, but these effects are negligible when $\Delta\tau < L^2$.

⁴This form is not symmetric with respect to R and R' . One can make better symmetric approximations, but the “primitive” form defined here is sufficient for convergence.

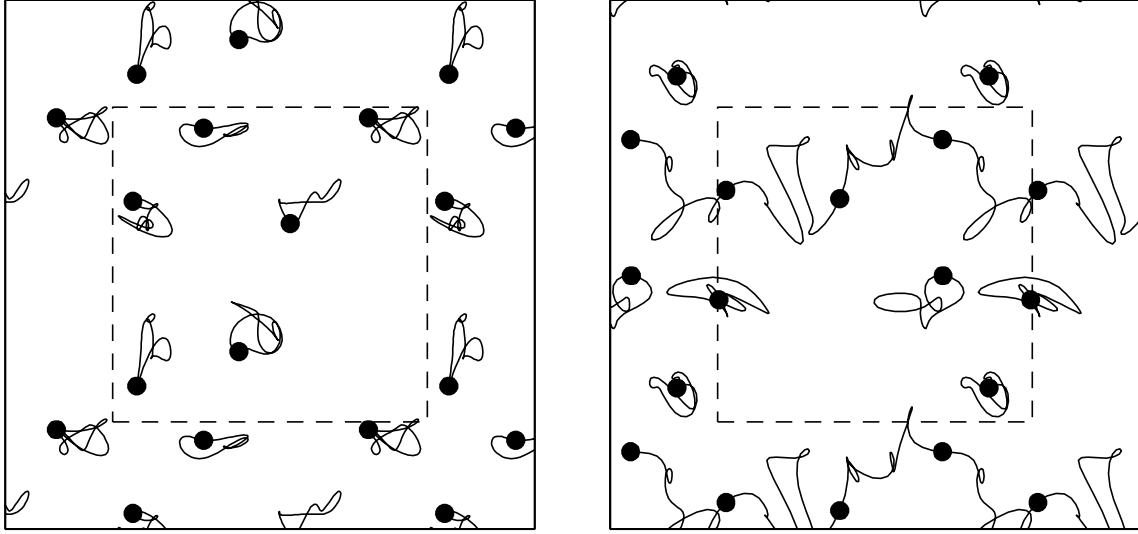


Fig. 1: Typical paths of six quantum particles in a 2D periodic square. The large black dots represent the positions of the particles at the start of their paths. The paths have been smoothed by zeroing their short wavelength Fourier components since a picture with all Fourier components would be a space-filling fractal curve, see ref. [3]. The left panel is the identity permutation, the right panel is a three particle permutation with a path “winding” in the horizontal direction across the periodic boundaries as indicated by the inner dotted square.

between neighboring beads (the second term in Eq. (8)) and an inter-polymer potential between different polymers (the third term in Eq. (8)). To calculate the partition function, and most thermodynamic properties, the polymers must close on themselves. The left panel of Fig. 1 shows a typical example of such paths. A lower temperature means a longer polymer, i.e., with more beads. Note that an individual polymer does not interact with itself except via the springs, and the inter-polymer potential (the third term in Eq. (8)) is not like for a real polymer: it only interacts with beads with the same path integral time-slice “index”.

To evaluate properties of the quantum system, we must perform the $3NM$ dimensional integral of Eq. (9) over all paths using either a generalized Metropolis Monte Carlo or molecular dynamics simulation. To obtain exact results within the statistical sampling error, we must calculate the results for several values of M and extrapolate to $M \rightarrow \infty$. In the following we use the notation for coordinates: $R_t = \{\mathbf{r}_{1,t}, \mathbf{r}_{2,t} \dots \mathbf{r}_{N,t}\}$: the first index is the particle index, the second index is the time-slice index.

For efficient computation, we need to improve the sampling and the action so as to reduce the needed number of time slices, as described in detail in Ref. [3]. To improve the action, it is advisable to use the pair action, i.e., the numerical solution to the 2-body problem [3]. For example, the divergence of the Coulomb potential when two unlike charges approach each other can wreak havoc on the stability of the algorithm since paths can fall into the region at small r_{ij} and never escape. A simple approach is to cut off the potential for $r_{ij} < \Delta\tau^{\frac{1}{2}}$, however, it is much better to use the exact two-body density matrix since it does not diverge as the charges approach each other, and its derivative obeys the cusp condition.

To compute the internal energy there are several approaches [3]. Differentiating the partition function with respect to temperature gives the *thermodynamic form*. The kinetic energy estimator is

$$K = \frac{3N}{2\Delta\tau} - \sum_{i,t} \left\langle \frac{(\mathbf{r}_{i,t} - \mathbf{r}_{i,t-1})^2}{4\lambda\Delta\tau^2} \right\rangle. \quad (10)$$

The second term is the negative spring energy of the classical polymer system but has the disadvantage that it diverges at small $\Delta\tau$ as $\Delta\tau^{-2}$, causing loss of efficiency if the time-step is extrapolated to zero. A form that has the same average value, the *virial* [4] estimator

$$K = \frac{3Nk_B T}{2} - \frac{1}{2} \left\langle \sum_{i=1}^N (\mathbf{r}_{i,t} - \bar{\mathbf{r}}_i) \cdot \mathbf{F}_{i,t} \right\rangle \quad (11)$$

does not have this problem. Here $\bar{\mathbf{r}}_i$ is the centroid position⁵ of particle i and $\mathbf{F}_{i,t}$ is the classical force on particle i at time t . Note that the first term is simply the classical kinetic energy. Quantum corrections are given in the second term and will vanish at high temperature since $|\mathbf{r}_{i,t} - \bar{\mathbf{r}}_i| \rightarrow 0$.

To perform an efficient sampling of Eq. (9) with Metropolis Monte Carlo, one needs to use collective moves [3] because single bead moves will make the whole procedure slow, i.e., require many steps to converge to the equilibrium distribution. For details on the path integral molecular dynamics approach, see [5].

2 Quantum statistics with path integrals

The most interesting consequences of quantum physics, e.g., superfluidity, Bose condensation, superconductivity, Fermi liquid behavior, come from the Fermi or Bose statistics of the particles. In the previous section, we did not consider particle statistics. The way to treat statistics is quite simple: we ignore the identities of particles so that when we close the paths in imaginary time, they can close on a permutation of themselves, i.e., $R_M = \hat{\mathcal{P}}R_0$ where $\hat{\mathcal{P}}R \equiv \{\mathbf{r}_{P_1}, \mathbf{r}_{P_2}, \dots, \mathbf{r}_{P_N}\}$ is a relabeling of the coordinates. To understand this pictorially, compare the left and right panels of Fig. 1.

To show that this procedure is correct, we first note that the wavefunctions of fermions (bosons) are antisymmetric (symmetric): their density matrix is defined by summing only over antisymmetric (symmetric) states⁶ in Eq. (1). In the following, we will denote the statistics of the particles by subscripts: ρ_F will denote the fermion density matrix, ρ_B the boson density matrix, ρ_D the Boltzmann (distinguishable particle) density matrix. The relabeling operator $\frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \hat{\mathcal{P}}$ projects out the states of correct symmetry. Here the upper sign (+1) is for bosons, and the lower sign (−1) is for fermions, where $(-1)^{\mathcal{P}}$ stands for the signature of the permutation: If a permutation is made of an odd number of pair exchanges it is negative, otherwise it is positive.

⁵The centroid is the center of mass of a given polymer, $\bar{\mathbf{r}}_i \equiv \beta^{-1} \int_0^\beta dt \mathbf{r}_{i,t}$. See [3] for the generalization to identical particles.

⁶A similar procedure can be used for other symmetries such as momentum or spin.

We use this operator to construct the path integral expression for bosons or fermions in terms of the Boltzmann density matrix

$$\rho_{B/F}(R, R'; \beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \rho_D(\hat{\mathcal{P}} R, R'; \beta). \quad (12)$$

Note that we could apply this relabeling operator to the first argument, the last argument or both; since the particles are identical the resulting density matrix would be the same. The connection between the Boltzmann density matrix and the bosonic or fermionic density matrix is important because it is the Boltzmann density matrix that arises naturally from paths. Including statistics, the path integral expression of the partition function becomes

$$Z_{B/F}(\beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \int dR_1 \dots dR_M \exp \left(- \sum_{t=1}^M S(R_{t-1}, R_t; \Delta\tau) \right) \quad (13)$$

with $R_0 = \hat{\mathcal{P}} R_M$.

2.1 Bosons

For bosons, the integrand in Eq. (13) is positive, but for large N it is very difficult to evaluate directly the permutation sum since it has $N!$ terms. However, we can enlarge the space to be sampled in the Monte Carlo random walk by including how the paths are connected, i.e., \mathcal{P} . One such connection is shown on the right panel of Fig. 1. With Monte Carlo techniques, this extra sampling does not necessarily slow down the calculation, but we need to include moves that are ergodic in the combined space of paths and connections as discussed in Ref. [3].

A macroscopic “percolation” of the polymers (i.e., a network of connected polymers spanning a macroscopic volume) is directly related to superfluidity [6]. Recall that any permutation can be decomposed into permutation cycles, i.e., into 2-, 3-, ... N -body exchange cycles. Superfluid behavior results when exchange cycles extending across a macroscopic distance appear at low temperature as we discuss now.

One of the fundamental properties of a Bose condensed system is superfluidity: a superfluid can flow without viscosity similar to how a superconductor can carry a current without resistance. The superfluid density is defined experimentally as follows: suppose the walls of a container are moved with a small velocity \mathbf{V} and the momentum acquired by the enclosed system in equilibrium is measured. In a normal liquid or solid, the enclosed system will move with the walls so that the acquired momentum will equal $M\mathbf{V}$ with M its total mass. However, a superfluid can shield itself from the walls. The superfluid fraction is defined in terms of the mass not contributing to the momentum:

$$\frac{\rho_s}{\rho} = 1 - \frac{\mathbf{P}}{M\mathbf{V}} \mapsto \frac{\langle \mathbf{W}^2 \rangle}{2\lambda\beta M}. \quad (14)$$

The expression on the right is how we calculate the superfluid fraction with imaginary-time path integrals in periodic boundary conditions [7]. In this expression we use “the winding number”

of a given path defined as

$$\mathbf{W} = \sum_{i=1}^N \frac{1}{\hbar} \int_0^\beta dt m_i \frac{d\mathbf{r}_{i,t}}{dt}. \quad (15)$$

It is the number of times a path wraps around the periodic boundaries in the x , y , or z directions. This remarkable formula relates the real-time linear response of moving the boundaries (or an impurity) to a topological property of imaginary-time path integrals. Since the size of a path of a single atom is its thermal de Broglie wavelength $\hbar\sqrt{mk_BT}$, which is always microscopic even at very low temperatures, the only way to have winding paths for a macroscopic cell, and, hence a non-zero superfluid density, is to have a permutation cycle that includes on the order of $N^{2/3}$ atoms (in 3D) such that the atoms, if linked together, can stretch across a macroscopic distance. There are interesting connections between the exchange of electrons in an insulator and the exchanges of bosonic paths which we will touch upon in the next section.

Bose condensation is another key property of superfluids that can be interpreted with path integrals. In a superfluid, a certain fraction of the particles will condense into the zero momentum state, or in an inhomogeneous system, into a single natural orbital. To determine the single particle density matrix, we need to sample paths where one particle does *not* close on itself; the two ends of an open “polymer” are free to move around the system. The single particle density matrix is defined as

$$n(\mathbf{r}, \mathbf{r}'; \beta) = \frac{1}{Z} \int d\mathbf{r}_2 \dots d\mathbf{r}_N \rho(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N, \mathbf{r}', \mathbf{r}_2, \dots, \mathbf{r}_N; \beta). \quad (16)$$

For a homogeneous system the momentum distribution is its Fourier transform:

$$n_{\mathbf{k}}(\beta) = \frac{1}{(2\pi)^3} \int d\mathbf{r} d\mathbf{r}' e^{-i\mathbf{k}(\mathbf{r}-\mathbf{r}')} n(\mathbf{r}, \mathbf{r}'; \beta). \quad (17)$$

For a normal, i.e., not Bose condensed, system the two ends in the single particle density matrix remain within a thermal de Broglie wavelength, implying that its Fourier transform, the momentum distribution is also localized. However, once macroscopic exchanges in the path can occur, the two ends can separate by a macroscopic distance so that $\lim_{|\mathbf{r}-\mathbf{r}'| \rightarrow \infty} n(\mathbf{r}, \mathbf{r}') \rightarrow n_0 > 0$ implying that $n_{\mathbf{k}} = n_0 \delta(k)$ where n_0 is the condensate fraction, the number of atoms with precisely zero momentum. The macroscopic exchange of particles is how the phase of the wavefunction is communicated.

Using Path-Integral Montecarlo (PIMC) one can calculate equilibrium properties of many-body ^4He at all temperatures both in the liquid phase above and below the superfluid transition, and in the solid phase. For details on the path integral theory of Bose superfluids and the PIMC calculations see ref. [3]. The worm algorithm [8] allows the sampling of a superfluid phase to be done more efficiently, particularly for systems with more than a few hundred bosons. It works in the grand canonical ensemble and can compute also unequal-time correlation functions such as the one particle Green's function in imaginary time.

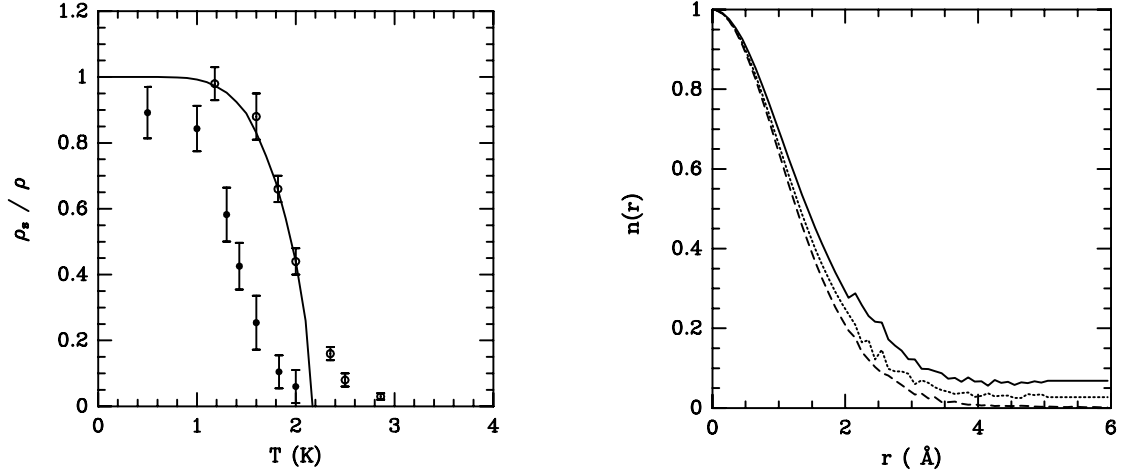


Fig. 2: Left panel: The ratio of superfluid density to total density: solid line, measured value for bulk ^4He at saturated vapour pressure; open circles, PIMC calculations with 64 atoms in PBC; solid circles, calculations for a droplet of 64 ^4He atoms. Right panel: The single-particle density matrix of ^4He above and below the lambda transition at temperatures 1.18, 2.22 and 4 K (from top to bottom).

2.2 Fermions

Suppose we do a path integral calculation of a fermion system by summing over permutations just as for bosonic systems but including the factor $(-1)^P$ as a weight in the numerator and denominator of any expectation value. If one performs an integration over a function having both positive and negative regions with Monte Carlo, the signal-to-noise-ratio, i.e., the efficiency, is much reduced. Doing a direct sampling of the boson paths and permutations and using the permutational sign to estimate properties of the fermion system leads [9] to a computational efficiency of the fermion system (ξ_F) that scales as

$$\xi_F = \xi_B e^{-2N\beta(\mu_F - \mu_B)}, \quad (18)$$

where μ_F (μ_B) is the free energy per particle of the fermion (boson) system and ξ_B the efficiency of the boson system. The direct fermion method, while exact, becomes exceedingly inefficient as $N\beta = N/k_B T$ increases – precisely when the physics becomes interesting.

2.3 Restricted path integral method

The restricted path identity (19) allows one to keep only “positive” paths at the cost of making an uncontrolled approximation. It is the generalization of the ground state fixed-node method: the nodes of the exact fermion density matrix give a rule for deciding which paths can contribute.⁷ The method is based on the identity

$$\rho_F(R_\beta, R_*; \beta) = \int dR_0 \rho_F(R_0, R_*; 0) \oint_{R_0 \rightarrow R_\beta \in \mathcal{Y}(R_*)} dR_t e^{-S[R_t]}, \quad (19)$$

⁷For this we need the path to be continuous. Lattice models or non-local Hamiltonians do not have continuous trajectories so this method is not as straightforward for those systems.

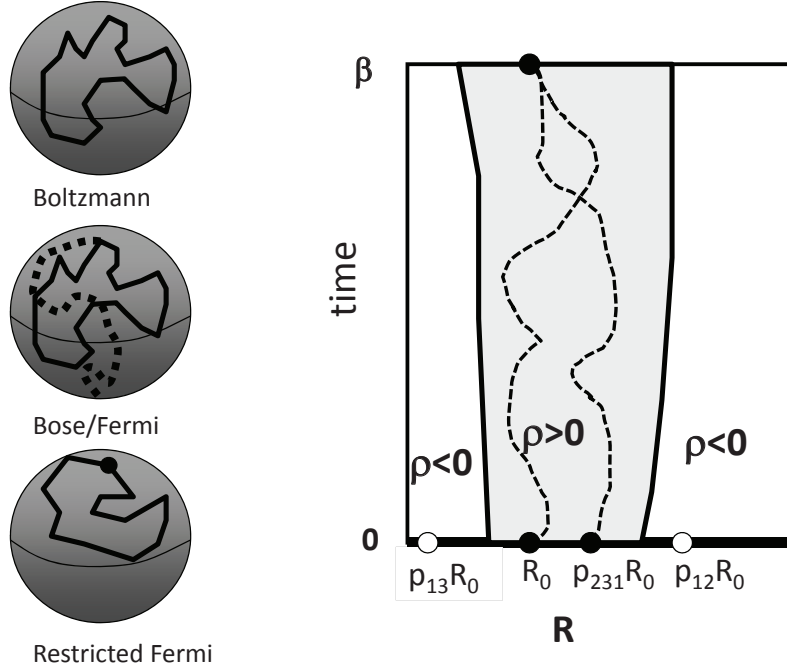


Fig. 3: Right panel: Space-time cartoon for proof of the restricted PIMC identity. The horizontal axis represents the spatial coordinates, the vertical axis imaginary time. The usual boundary conditions are delta functions at $t = 0$ represented as dots on the horizontal axis; positive values for even permutations (full circles) and negative values for odd permutations (open circles). However, it is sufficient to work in the strictly positive domain (shaded), if the nodal domain is correct. Shown are two allowed paths in this domain, one from the identity permutation, the other from a 3-body permutation. Left panel: Depiction of path integrals for distinguishable ions (top), ortho- and para- hydrogen (middle) and restricted paths for ortho-hydrogen (bottom); its reference point is the large dot.

where the subscript on the path integration means that we integrate only over paths that start at R_0 , end at R_β and are node-avoiding, i.e., for which $\rho_F(R_t, R_*; t) \neq 0$ for all $0 < t < \beta$; here the “reference point” R_* defines the nodes. To prove this identity, we note that the fermion density matrix satisfies the Bloch Equation

$$\frac{\partial \rho_F(R; \beta)}{\partial \beta} = \lambda \sum_{i=1}^N \nabla_i^2 \rho_F(R; \beta) - V(R) \rho_F(R; \beta) \quad (20)$$

with the initial conditions

$$\rho_F(R, R_*; 0) = \frac{1}{N!} \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \delta(R - \hat{\mathcal{P}} R_*). \quad (21)$$

Hence, the path starts at a permutation, \mathcal{P} , of the reference point, $R_0 = \hat{\mathcal{P}} R_*$ and carries a weight $\frac{1}{N!} (-1)^{\mathcal{P}}$. The solution of the Bloch equation is uniquely specified by its boundary conditions, just like the Poisson equation in electrostatics [10]. Normally, one uses the values at zero imaginary time, i.e., infinite temperature, as boundary conditions. We can, however, also take the nodal surfaces: $\rho_F(R, R_*; \beta) = 0$ as boundary conditions as illustrated in Fig. 3: we

want the solution of the Bloch equation that vanishes on a preselected nodal surface. We enforce this solution by putting an infinite repulsive wall on this surface, or, equivalently, restricting the allowed paths to remain on the interior of a given nodal domain, $\mathcal{V}(R_*)$. The solution is exact if the assumed nodes are correct. For the diagonal elements of the density matrix, $R_\beta = R_*$, the contributions of all paths must be positive, hence, the sum over permutations is restricted to even permutations.⁸

Calculations with restricted paths have been done on a variety of simple fermion systems. Recently the energy of the homogenous electron gas throughout its phase diagram has been determined. Earlier calculations have been performed [11] on liquid ^3He and hot dense hydrogen [12–15].

The restricted path picture is a novel way of analyzing fermion systems [9, 11]. First consider a Fermi liquid. That a Fermi liquid has exchange paths can be understood by considering its momentum distribution, n_k : By definition a Fermi liquid has a discontinuity in n_k at k_F . Using properties of Fourier transforms, this implies $\rho(\mathbf{r}, \mathbf{r}') \propto |\mathbf{r} - \mathbf{r}'|^{-3}$ at large separations. Such a slow decay can only come from macroscopic exchanges of even permutations. In a superconductor with Cooper pairs of electrons, there will be paired up-spin and down-spin macroscopic exchanges [16]. Krüger and Zaanen [17] have interpreted other quantum phase transitions in terms of the restricted path formalism.

The problem we now face for calculation is that the unknown fermion density matrix appears on the right-hand side of Eq. (19), since it is used to define the criterion of node-avoiding, as well as the left-hand side. To apply the formula directly, we would have to self-consistently determine the nodes. In practical calculations, we make an ansatz for the nodal surfaces, such as using the nodes of the density matrix from a mean-field theory.

The reference point, R_* plays a very special role in restricted path integrals since it restricts the paths as illustrated in the example below. For boson or distinguishable particle path integrals, all time slices are equivalent, but restricted paths break this time symmetry. For fermions we can use a “ground-state” restriction that does not depend on the reference point. This can be achieved by using an antisymmetric trial wavefunction $\Psi(R)$ and requiring that $\Psi(R_t) \neq 0$ throughout the path.

3 Exchange of localized particles

We now discuss a specialized application of PIMC, namely the computation of exchange frequencies between electrons localized on different lattice sites. First we discuss a simple model: we confine a single electron to the interior of the union of two spheres as shown in Fig. 4. Because of mirror symmetry, the quantum states can be classified by parity. The splitting between

⁸ We have done more than simply restricting the sum over permutations to even permutations. We only take those even permutations that also stay in the nodal domain. The reason that restriction gives the same result is that negative paths can be paired with positive paths and canceled. The gradient of the density matrix at the node is the flux of path and since the gradient is continuous across the node, the positive paths crossing at a given nodal point will precisely cancel against the negative paths.

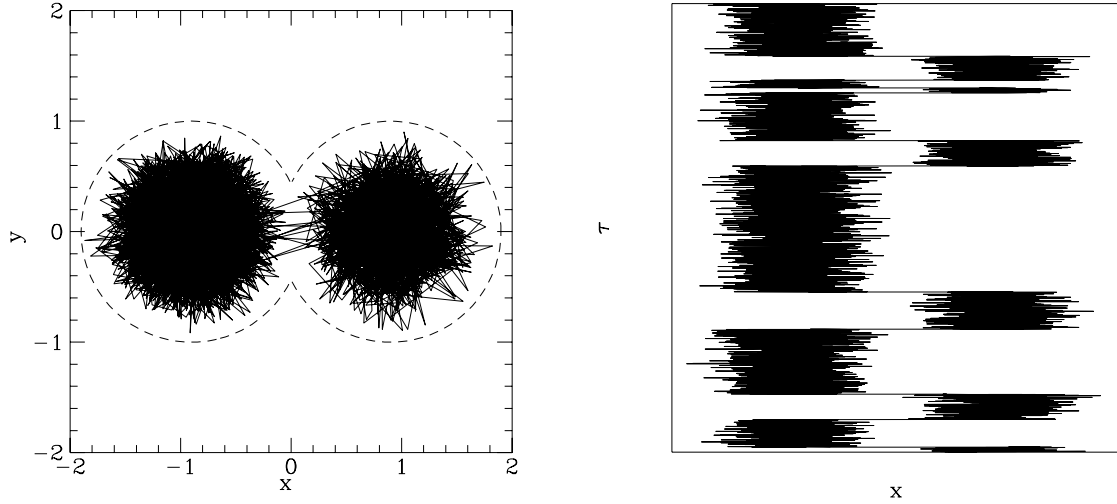


Fig. 4: *Left panel: The imaginary time path confined to the interior of two spheres (shown with dashed lines). Right panel: the same path shown as a function of imaginary time. The electron stays inside one sphere for a long period, until it finds the duct to the other sphere.*

the lowest even and odd parity states defines the exchange frequency, $J = (E_1 - E_0)/2 > 0$. A wavefunction, initially localized in one of the spheres, will oscillate back and forth with a frequency given by J/\hbar . Let us suppose that the splitting energy is much less than the zero-point energy, so that higher excitations can be neglected.

Here we show how to calculate this frequency using path integrals. In Fig. 4 we show an example of a world-line diagram of the imaginary-time paths in the double-sphere model. We see that the electron spends a long time in a single sphere, but occasionally “tunnels”⁹ over to the other sphere. The tunneling is rapid, since the wavefunction is squeezed as it passes from one sphere to the other, costing energy.

Let us denote the coordinates of the centers of the two spheres as Z and $\hat{P}Z$; the motivation for this notation will become clear when we discuss the multi-electron generalization. Now define f_P as the ratio of the imaginary-time matrix element connecting Z to $\hat{P}Z$ with that connecting Z to itself:

$$f_P(\beta) = \frac{\rho_D(Z, \hat{P}Z; \beta)}{\rho_D(Z, Z; \beta)}. \quad (22)$$

If we now assume that β is large enough that only the lowest two states contribute to the density matrix, then:

$$f_P(\beta) = \begin{cases} 0 & \text{if } \beta < \beta_0, \\ \tanh(J(\beta - \beta_0)) & \text{if } \beta > \beta_0. \end{cases} \quad (23)$$

Here $\beta_0 = \ln(\varphi_1(Z)/\varphi_0(Z))/J$ with φ_1 and φ_0 being the eigenstates corresponding to energies E_1 and E_0 . The rate in imaginary time (in units of \hbar) for the electron to cross from one sphere to the other is J . In the polymer (imaginary time path integrals) language, J is related to the free energy it takes to pull a single end of a “linear polymer” from one sphere to the other and can be estimated with special techniques [18, 19].

⁹Tunnels is in quotes because we are in imaginary time, not real time. The imaginary-time transversal of the barrier is called an “instanton” because it takes place so quickly.

Now, let us generalize from the two-sphere model to a many-body system. We follow the theory of Thouless [20], based on the earlier work of Dirac [21] on electronic exchange. Consider a system of spinless electrons in a perfect crystal when the electrons are localized with a single electron per unit cell. Because exchange is rare for localized electrons, we can label the positions of electrons in a crystalline lattice by localized Wannier orbitals: Z denotes the N -body coordinates resulting in one such assignment of N electrons to N orbitals, $\hat{P}Z$ the effect of applying the permutation \hat{P} to that assignment. If there are N electrons and N Wannier functions, there are $N!$ such assignments, so there is an $N!$ degeneracy of the ground state in the absence of electron exchange. The splitting induced by tunneling between states Z and $\hat{P}Z$ is defined to be $2J_P$ as in the example above. All of the previous discussion concerning how to calculate J_P with path integrals, Eqs. (22) and (23), then applies.

Following Dirac and Heisenberg, electronic exchange couples the electron spins on different atoms and for pair exchange results in a Heisenberg spin Hamiltonian

$$\hat{\mathcal{H}} = - \sum_{\mathcal{P}} J_P (-1)^{\mathcal{P}} \hat{P}_{\sigma} = -J_2 \sum_{(i,j)} \sigma_i \cdot \sigma_j, \quad (24)$$

where in the first summation \mathcal{P} ranges over all $N!$ permutations,¹⁰ $(-1)^{\mathcal{P}}$ is the sign of the permutation, and \hat{P}_{σ} permutes spins. With this argument Thouless [20] showed that exchange of an even number of spins favors antiferromagnetism while exchange of an odd number of spins favors ferromagnetism. The second equation, the conventional Heisenberg Hamiltonian, applies if the only exchanges allowed involve two- and three- body permutations. A clear discussion is given by Roger [22].

PIMC calculations have been used to determine the Heisenberg exchange coefficients in the Wigner crystals [23, 24] and in solid ^3He . The PIMC method to determine the exchange frequencies is much superior to one based on Projector Monte Carlo, i.e., Diffusion Monte Carlo, since one can determine directly the terms in the underlying spin Hamiltonian and the results are accurate even if the exchange frequencies are very small. Calculations of the exchange frequencies of the 2D Wigner crystal suggest a frustrated spin liquid phase may be stable [23]. These methods have not yet been applied to realistic electronic materials.

The methods have been applied extensively to solid ^3He which forms a *bcc* crystal. If pair exchanges dominated, the bipartite lattice structure would order into an antiferromagnetic state. However, experimentally the ground state is found to be in a symmetry-broken spin state with 8 atoms per unit cell. Using PIMC, we found that this structure results from a competition between even and odd ring exchanges. As the density of the crystal is lowered near to the melting density, it is found that long exchanges become probable (cycles of up to 10 atoms were considered [25]). This suggests a picture for how a metal/insulator transition could occur: as a localized system gets near the metal-insulator transition, the energy to create a vacancy-interstitial pair goes to zero, and longer and longer ring exchange cycles become important. Once the transition occurs, this picture of ring exchanges breaks down.

¹⁰One need only consider cyclic permutations of neighboring electrons, otherwise $J_{\mathcal{P}}$ will be much smaller. Hence we need only consider ring exchanges.

4 PIMC calculations of supersolid helium

Recent torsional-oscillator observations by E. Kim and M.H.W. Chan on solid ^4He [26], have revived interest in the supersolid phase. In this phase, one has both long-range crystalline order and superfluidity. Using PIMC, we examined [27] whether crystalline *hcp* helium, assumed to be free of defects such as impurities and vacancies, could have a supersolid phase. One might think that there would always be ground state defects, arising from the large quantum zero point fluctuations. Near melting, the root-mean-square vibration about the lattice site is 30%, so that at any instant of time, a good fraction of atoms are closer to a neighboring site than to their home site. However, the absence of an atom from a lattice site is not sufficient for having a supersolid; if the empty site is always accompanied by a nearby doubly occupied site, there will be no mass current.

Path integrals give a much cleaner framework for determining whether bulk helium could be supersolid. They can be used to compute the superfluid density and the momentum distribution without the assumption of a trial wave function or any other uncontrolled approximation. Using methods [18, 19] developed for solid ^3He described above, we calculated the exchange frequencies for solid ^4He and estimated how close they were to the critical value for superfluidity. The frequencies for 2, 3, and 4 atom exchanges were very small [3], e.g., $J_2 \sim 3\mu K$ at melting density. However, small cyclic exchanges are quite different from the long exchanges needed to get a supersolid. Fig. 5 shows the results of calculations of the frequency of the simplest straight line winding exchanges in the basal plane of the *hcp* crystal. We found that the exchange frequencies decreased exponentially with the number of atoms in the exchange. Using a model for all exchanges we concluded that in solid ^4He only localized exchanges will be present and thus it should not exhibit the property of nonclassical rotational inertia. We also computed the single particle density matrix from Eq. 16 (see Fig. 5) and found that, since it goes exponentially to zero at large separation, the condensate fraction will vanish. Thus, based on other PIMC calculations, we think it unlikely that the observed phenomena of Kim and Chan are due to vacancies or ^3He impurities. Recent experiments have confirmed these computational findings.

5 Lexicon of the quantum-classical isomorphism

As we have mentioned earlier, there is an exact correspondance between quantum statistical mechanics and the classical statistical mechanics of imaginary time path integrals. There is an exact, systematic procedure for understanding many properties of quantum systems purely in terms of classical statistical mechanics. Note that there is a curious shift of vocabulary in going from the quantum system to the polymer model. Scientists discussing path integrals sometimes resemble children playing the game of “opposites,” where the child says the opposite of what is intended (“I do not want a cookie.”) Usually the game quickly degenerates into confusion because common language is ambiguous and not entirely logical as the opposite of a given statement is non-unique. Discussions of path integrals should be clearer since path integrals are based on mathematics, but the translation is complicated by several features.

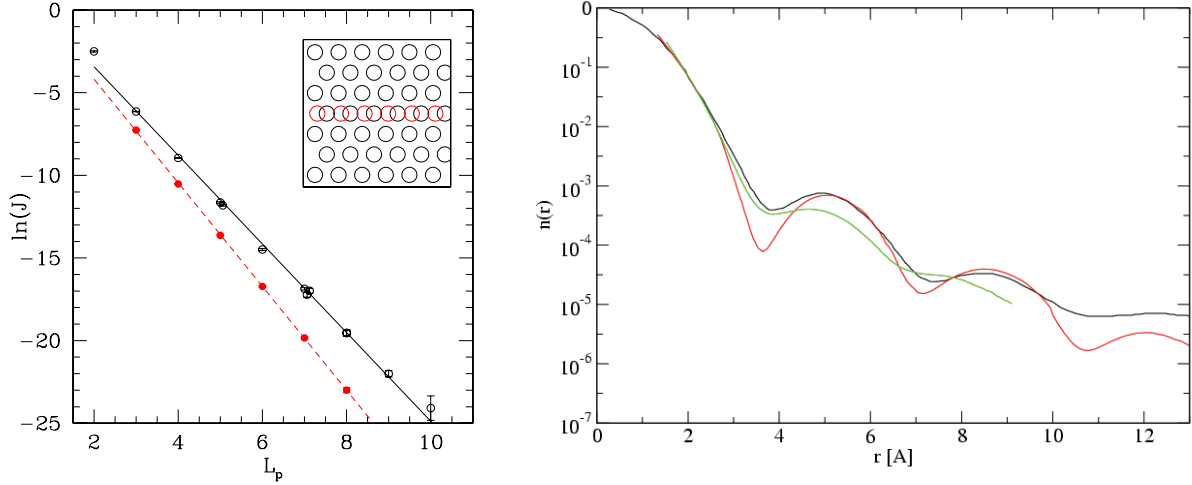


Fig. 5: Left panel: The exchange frequencies J in K versus the exchange length L_p for straight line exchanges in the basal plane that wind around the periodic cell (see inset) [27]. Right panel: Single particle density matrix in solid helium [28].

The same word applied to the quantum system and the classical system can mean quite different things, e.g., it is confusing to refer to the “energy” of the polymer model or its entropy. The entropy of a quantum system decreases with temperature but at low temperature, the corresponding polymer system becomes more disordered. The confusion arises because the “temperature” of the polymer model is not equal to the quantum temperature. To translate what we mean by temperature into the polymer model we must find how β appears in the action.¹¹ The lower the temperature, the more beads are on the polymer. Zero temperature corresponds to infinitely long chains.

Time is a word that can have at least three different meanings: real time in the quantum system, the “imaginary time” of the path integrals, or the time related to how the path is moved in the computer program. If we confuse the first two meanings of time, a word can have exactly the opposite meaning in the quantum and polymer systems. For example, the “velocity” of a bead is usefully defined as its displacement from one time slice to the next, divided by τ . But with this definition atoms that are “fast” correspond to low-energy atoms because they are spread out and their kinetic energy is small. On the other hand, particles that are trapped in a small region have a small “velocity” and a high energy. It is possible for a single realization of a path to have a negative kinetic energy by being spread out more than usual, but the average over all paths must be positive. The inversion of meaning comes because path integrals are in imaginary time. Any observable corresponding to a scalar function of coordinates maps trivially from the quantum system into the polymer model. For example, the particle density is simply the average density of the beads

$$\langle \rho(\mathbf{r}) \rangle = \left\langle \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_{im}) \right\rangle. \quad (25)$$

¹¹It is best not to see how the time step appears in the action because the time step is fixed by requiring that the action be accurate. Hence the spring constant and the interbead potential should be fixed as temperature varies. This means that β will be proportional to the number of time slices.

There are often several different ways of mapping a quantum concept onto the classical system. A concept such as superfluidity is very general and related to many quantum-mechanical observables. In a few words, superfluidity is equivalent to the presence of macroscopic polymers in the classical model.

This aspect of figuring out different ways of calculating quantum properties in some ways resembles experimental physics. The theoretical concept may be perfectly well defined, but it is up to the ingenuity of the experimentalist to find the best way of doing the measurement. Even what is meant by “best” is subject to debate. Many important quantities of quantum systems are really defined as dynamical quantities, while the quantum-classical correspondence is restricted to imaginary time. Often, one can reformulate the quantum property in imaginary time, but not always. There is still much to be done in learning how to exploit the quantum-classical correspondence. To conclude, we summarize the relationship between quantum concepts and the classical polymer language with the following lexicon:

Bose condensation : delocalization of ends of an open polymer

Boson statistics : allowing the possibility that polymers can hook up in any possible way

Cooper pairing : paired fermion (restricted) polymers

degeneracy temperature : a condition in which polymers are dense enough and extended enough that they touch and can exchange

density : the bead density

exchange energy : logarithm of the fraction of monomers (times $k_B T$)

exchange frequency in a crystal : free energy to link polymers in a polymer crystal

Fermi liquid : winding fermion (restricted) polymers

free energy : free energy of a system of ring polymers

imaginary velocity : bond vector

insulator : localized exchanging polymers

kinetic energy : negative spring energy

moment of inertia : the mean-squared area of ring polymers

momentum correlation function : bond-bond correlation

momentum distribution : Fourier transform of end-end distribution

pair-correlation function : pair-correlation function between beads at the same “time”

Pauli principle : restricted polymers

particle : ring polymer

potential energy : iso-“time” potential between beads

single particle density matrix : the end-to-end distribution of an open polymer

superfluid density : the mean-squared winding number

superfluid state : a state in which a finite fraction of polymers are hooked together in polymers of macroscopic size

temperature :

1. inverse polymer length,
2. inverse coupling constant for the inter-polymer potential,
3. spring constant between neighboring beads

thermal wavelength : polymer extension

References

- [1] R.M. Martin, L. Reining, and D.M. Ceperley: *Interacting Electrons* (Cambridge University Press)
- [2] H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959)
- [3] D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995)
- [4] M.F. Herman, E.J. Bruskin, and B.J. Berne, J. Chem. Phys. **76**, 5150 (1982)
- [5] M.E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford University Press, Oxford, 2010)
- [6] R.P. Feynman, Phys. Rev. **91**, 1291 (1953)
- [7] E.L. Pollock and D.M. Ceperley, Phys. Rev. B **36**, 8343 (1987)
- [8] M. Boninsegni, N.V. Prokof'ev, and B.V. Svistunov, Phys. Rev. E **74**, 036701 (2006)
- [9] D.M. Ceperley: *Path integral Monte Carlo methods for Fermions* in K. Binder and G. Ciccotti (eds.):
Monte Carlo and Molecular Dynamics of Condensed Matter Systems (Editrice Compositori, Bologna, 1996)
- [10] J.D. Jackson: *Classical Electrodynamics* (Wiley, New York, 1962)
- [11] D.M. Ceperley, Phys. Rev. Lett. **69**, 331 (1992)
- [12] C. Pierleoni, B. Bernu, D.M. Ceperley, and W.R. Magro, Phys. Rev. Lett. **73**, 2145 (1994)
- [13] W.R. Magro, D.M. Ceperley, C. Pierleoni, and B. Bernu, Phys. Rev. Lett. **76**, 1240 (1996)
- [14] B. Militzer and D.M. Ceperley, Phys. Rev. Lett. **85**, 1890 (2000)
- [15] B. Militzer and D.M. Ceperley, Phys. Rev. E **63**, 66404 (2001)
- [16] M. Boninsegni and D.M. Ceperley, J. Low Temp. Phys. **104**, 339 (1996)
- [17] F. Krüger and J. Zaanen, Phys. Rev. B **78**, 035104 (2008)
- [18] D.M. Ceperley and G. Jacucci, Phys. Rev. Lett. **58**, 1648 (1987)
- [19] B. Bernu and D.M. Ceperley: *Calculations of exchange frequencies with path integral Monte Carlo: solid ^3He adsorbed on graphite* in M.P. Nightingale and C.J. Umrigar (eds.):
Quantum Monte Carlo Methods in Physics and Chemistry (Kluwer, 1999), pp. 161–182
- [20] D.J. Thouless, Proc. Phys. Soc. London **86**, 893 (1965)

- [21] P.A.M. Dirac, Proc. Roy. Soc. London **Ser. A 123**, 714 (1929)
- [22] M. Roger, J. Low Temp. Phys. **162**, 625 (2011)
- [23] B. Bernu, L. Candido, and D.M. Ceperley, Phys. Rev. Lett. **86**, 870 (2001)
- [24] L. Candido, B. Bernu, and D.M. Ceperley, Phys. Rev. B **70**, 094413 (2004)
- [25] L. Cândido, G.-Q. Hai, and D.M. Ceperley, Phys. Rev. B **84**, 064515 (2011)
- [26] E. Kim and M.H.W. Chan, Science **305**, 1941 (2004)
- [27] D.M. Ceperley and B. Bernu, Phys. Rev. Lett. **93**, 155303 (2004)
- [28] B.K. Clark and D.M. Ceperley, Phys. Rev. Lett. **96**, 105302 (2006)

15 Auxiliary-Field Quantum Monte Carlo for Correlated Electron Systems

Shiwei Zhang

College of William & Mary

Department of Physics, Williamsburg, VA23185, USA

Contents

1	Introduction	2
2	Formalism	4
2.1	Ground-state projection	4
2.2	Slater determinant space	6
2.3	Hubbard-Stratonovich transformation	8
2.4	A simple example	10
3	Ground-State AFQMC Methods	12
3.1	Free-projection AFQMC	12
3.2	Why and how does the sign problem occur?	13
3.3	The constrained path Monte Carlo method	16
3.4	The phaseless formalism for complex auxiliary-fields	18
4	Illustrative results	23
5	Concluding remarks	25
6	Acknowledgments	26
A	A few basics of Monte Carlo techniques	27

1 Introduction

Predicting materials properties requires robust and reliable calculations at the most fundamental level. Often the effects being studied or designed originate from electron correlations, and small errors in their treatment can result in crucial and qualitative differences in the properties. Accurate treatment of electron correlations and collective behaviors is one of the great challenges in modern science [1, 2]. Explicit solution of the many-body Schrödinger equation leads to rapidly growing computational cost as a function of system size: exponential, as in configuration interaction (CI) [3], or at the minimum, a high power, e.g., $\mathcal{O}(N^7)$ as in the coupled-cluster CCSD(T) [3], the preeminent quantum chemistry method.

To circumvent the problem, most computational quantum mechanical studies of large, realistic systems rely on simpler independent-particle approaches based on density-functional theory (DFT) [2, 4], using an approximate energy functional to include many-body effects. These replace the electron-electron interaction by an effective potential, thereby reducing the problem to a set of one-electron equations. These methods have often been extremely effective in complex molecules and solids [4], and are the standard in electronic structure, widely applied in condensed matter, quantum chemistry, and materials science.

Despite the tremendous successes of DFT (most noticeably through its Car-Parrinello molecular dynamics implementation [5]), the treatment of electronic correlation is approximate. For strongly correlated systems (e.g., high-temperature superconductors, heavy-fermion metals, magnetic materials, optical lattices), where correlation effects from particle interaction crucially modify materials properties [6–9], the approximation can lead to qualitatively incorrect results. Even in moderately correlated systems when the method is qualitatively correct, the results are sometimes not sufficiently accurate. For example, in ferroelectric materials the usually acceptable 1% errors in DFT predictions of the equilibrium lattice constant can lead to almost full suppression of the ferroelectric order.

The development of alternatives to independent-particle theories is therefore of paramount fundamental and practical significance. Quantum Monte Carlo (QMC) methods [10–15] are among the most promising candidates for post-mean-field calculations. By using stochastic sampling rather than explicit integration over phase space, these methods have computational costs that scale gracefully (N^3 – N^4) with system size and exceptional potential for parallel scaling [16].

For fermion systems, however, QMC methods suffer from the so-called “sign” problem [17–19]. In these systems, the Pauli exclusion principle requires that the states be anti-symmetric under interchange of two particles. As a consequence, negative signs appear, which cause cancellations among contributions of the Monte Carlo (MC) samples of phase space. In fact, as the temperature is lowered or the system size is increased, such cancellation becomes more and more complete. The net signal thus decays *exponentially*. The algebraic scaling is then lost, and the method breaks down. Clearly the impact of this problem on the study of correlated electron systems is extremely severe.

To date most applications of QMC methods to correlated electron systems have either lived with the sign problem or relied on some form of approximation to overcome the exponential

scaling. The former is seen more often in the community studying lattice models for strongly correlated systems where, understandably, it is more difficult to make approximations and more demanding of insensitivity of the many-body method to *a priori* input. However, this approach clearly has severe limitations in the systems that can be examined. The latter has been more prevalent in electronic structure, where without some control of the sign problem it is difficult to sustain signal in even the simplest realistic Hamiltonian. This approach loses exactness and the results can be biased by the approximation.

The approach we take in these lecture notes is in some sense to combine the two approaches. The sign/phase problem will be controlled approximately. We cast the MC random walks in a space of over-complete Slater determinants, which significantly reduces the severity of the sign problem. In this space we formulate constraints on the random-walk paths which lead to better approximations that are less sensitive to the details of the constraint. We then develop internal checks and constraint release methods to systematically improve the approach.

These methods have come under the name of constrained path Monte Carlo (CPMC) [20] for systems where there is a sign problem (for example, Hubbard-like models where the local interactions lead to auxiliary-fields that are real). For electronic systems where there is a phase problem (as the Coloumb interaction leads to complex fields), the methods have been referred to as phaseless or phase-free auxiliary-field QMC (AFQMC) [15, 21]. In both cases, the idea is to constrain the sign or phase of the overlap of the sampled Slater determinants with a trial wave function. It eliminates the sign or phase instability and restores low-power (typically to the third power of system size) computational scaling. Applications to a variety of systems have shown that the methods are very accurate, even with simple trial wave functions taken directly from mean-field calculations (see, e.g., Refs [22] and [23] references therein). As an example, in Ref. [23], an 8×128 lattice in a doped repulsive Hubbard model (with over 900 electrons) is studied to calculate the ground-state magnetic and charge-correlations. The dimension of the Hilbert space for this system is $\mathcal{O}(10^{600})!$

With these notes, we will give a pedagogical introduction to the AFQMC method, starting from the forms without any constraint [13, 14], then discussing the sign and phase problem [18, 19, 24], and ideas for constraining the paths in the path integral in AF space [15, 20, 25] to eliminate the problem. The chapter is not meant to be a comprehensive review. Rather, we focus on several basic algorithmic aspects, given the length restrictions of the chapter, to complement the lectures. We will cover ground-state calculations in Hubbard-like models [19, 20, 26] and in realistic materials [15, 21, 27]. We will not cover finite-temperature calculations, although corresponding methods have been developed [28, 29], nor will we discuss boson or mixture systems, although the algorithms share many common features and some of the papers on boson systems [29–31] contain details that can serve as useful references. Some of the more pedagogical materials in this chapter are based on Refs. [19] and [24]. The references cited are sources where we have drawn materials for the lectures, and point to possible additional reading; they are by no means a literature survey.

2 Formalism

2.1 Ground-state projection

The Hamiltonian for any many-fermion system with two-body interactions (e.g., the electronic Hamiltonian under the Born-Oppenheimer approximation) can be written as

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = -\frac{\hbar^2}{2m} \sum_{m=1}^M \nabla_m^2 + \sum_{m=1}^M V_{\text{ext}}(\mathbf{r}_m) + \sum_{m<n}^M V_{\text{int}}(\mathbf{r}_m - \mathbf{r}_n) \quad (1)$$

where \mathbf{r}_m is the real-space coordinate of the m^{th} fermion, $V_{\text{ext}}(\mathbf{r}_m)$ is the effect of the external potential on it, and $V_{\text{int}}(\mathbf{r}_m - \mathbf{r}_n)$ is the interaction between fermions m and n . We have represented the external potential as local, although this does not have to be the case. For example, in plane-wave calculations we will use a norm-conserving pseudopotential. This will lead to a non-local function V_{ext} [27]. The total number of fermions, M , will be fixed in the calculations we discuss. For simplicity, we have suppressed spin-indices, but the spin will be made explicit when necessary. In that case, M_σ is the number of electrons with spin σ ($\sigma = \uparrow$ or \downarrow). We assume that the interaction is spin-independent, so the total S_z , defined by $(M_\uparrow - M_\downarrow)$, is fixed in the calculation, although it will be straightforward to generalize our discussions to treat other cases, for example, when there is spin-orbit coupling.

With any chosen one-particle basis, for example lattice sites (Hubbard model), plane-waves (as in solid state calculations) [27], or Gaussians (as in quantum chemistry) [21, 32], the Hamiltonian can be written in the general form

$$\hat{H} = \hat{H}_1 + \hat{H}_2 = \sum_{i,j}^N T_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{i,j,k,l}^N V_{ijkl} c_i^\dagger c_j^\dagger c_k c_l, \quad (2)$$

where N is the size of the chosen one-particle basis, and c_i^\dagger and c_i are the corresponding creation and annihilation operators. The one-body matrix elements, T_{ij} , contain the effect of both the kinetic energy and external potential, while the two-body matrix elements, V_{ijkl} , are from the interaction. The matrix elements are integrals expressed in terms of the basis functions and the potentials. We will assume that they have been evaluated and are known as we begin our many-body calculations. Examples are given below.

One of the simplest Hamiltonians of this form is the Hubbard model, which has played an important role in many-body physics and whose properties are still not fully understood:

$$\hat{H} = \hat{K} + \hat{V} = -t \sum_{\langle i,j \rangle \sigma}^N c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i^N n_{i\uparrow} n_{i\downarrow}. \quad (3)$$

Here N is the number of lattice sites, $c_{i\sigma}^\dagger$ and $c_{j\sigma}$ are creation and annihilation operators of an electron of spin σ on the i -th lattice site, t is the nearest-neighbor hopping energy, $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ is the density operator, and U is the interaction strength. Two parameters, the interaction U/t and the electron density $(M_\uparrow + M_\downarrow)/N$, determine the physics given the topology of the lattice.

Most ground-state QMC methods are based on

$$|\Psi_0\rangle \propto \lim_{\tau \rightarrow \infty} e^{-\tau \hat{H}} |\Psi_T\rangle; \quad (4)$$

that is, the ground state $|\Psi_0\rangle$ of a many-body Hamiltonian \hat{H} can be projected from any known trial state $|\Psi_T\rangle$ that satisfies $\langle \Psi_T | \Psi_0 \rangle \neq 0$. In a numerical method, the limit can be obtained iteratively by

$$|\Psi^{(n+1)}\rangle = e^{-\Delta\tau \hat{H}} |\Psi^{(n)}\rangle, \quad (5)$$

where $|\Psi^{(0)}\rangle = |\Psi_T\rangle$. Ground-state expectation $\langle \hat{O} \rangle$ of a physical observable \hat{O} is given by

$$\langle \hat{O} \rangle = \lim_{n \rightarrow \infty} \frac{\langle \Psi^{(n)} | \hat{O} | \Psi^{(n)} \rangle}{\langle \Psi^{(n)} | \Psi^{(n)} \rangle}. \quad (6)$$

For example, the ground-state energy can be obtained by letting $\hat{O} = \hat{H}$. A so-called mixed estimator exists, however, which is exact for the energy (or any other \hat{O} that commutes with \hat{H}) and can lead to considerable simplifications in practice:

$$E_0 = \lim_{n \rightarrow \infty} \frac{\langle \Psi_T | \hat{H} | \Psi^{(n)} \rangle}{\langle \Psi_T | \Psi^{(n)} \rangle}. \quad (7)$$

QMC methods carry out the iteration in Eq. (5) by Monte Carlo sampling. The difference between different classes of methods amounts primarily to the space that is used to represent the wave function or density matrix and to carry out the integration. The auxiliary-field QMC (AFQMC) methods work in second quantized representation and in an auxiliary-field space, while Green's function Monte Carlo (GFMC) or diffusion Monte Carlo (DMC) works in first-quantized representation and in configuration space [10, 11].

Let us assume that $|\Psi_T\rangle$ is of the form of a single Slater determinant or a linear combination of Slater determinants. A Slater determinant is the form of a mean-field solution to \hat{H} (see next section) expressed in terms of the basis functions chosen in Eq. (2). The operation of $e^{-\tau \hat{H}_1}$ on a Slater determinant is straightforward to calculate, and it simply yields another determinant. The ground-state projection would therefore turn into the propagation of a single Slater determinant if it were somehow possible to write the two-body propagator $e^{-\tau \hat{H}_2}$ as the exponential of a one-body operator. This is realized in mean-field methods. In the Hartree-Fock (HF) approximation \hat{H}_2 is replaced by one-body operators times expectations with respect to the current Slater determinant wave function, schematically:

$$c_i^\dagger c_j^\dagger c_k c_l \rightarrow c_i^\dagger c_l \langle c_j^\dagger c_k \rangle - c_i^\dagger c_k \langle c_j^\dagger c_l \rangle. \quad (8)$$

A decomposition that includes pairing is also possible, leading to a Hartree-Fock-Bogoliubov calculation. In the local density approximation (LDA) in DFT [4], \hat{H}_2 is replaced by the density operator in real-space times a functional of the local density calculated with respect to the current Slater determinant in the self-consistent process. In both these cases, an iterative procedure can be used, following Eq. (5), to project out the solution to the approximate Hamiltonians, as an imaginary-time evolution of a single Slater determinant [33].

2.2 Slater determinant space

The building blocks of Slater determinants are single-particle basis states. These single-particle basis states can be plane waves, or lattice sites in the Hubbard model, Gaussians, or energy eigenstates in a mean-field potential. Often the space of single-particle basis states is truncated. Single-particle wave functions (orbitals) are formed with the basis states. Slater determinants are then built from the single-particle orbitals.

We define some notations that we will use throughout the discussion [24]:

- N : number of single-electron basis states. For example, N can be the number of lattice sites ($L \times L$) in the two-dimensional Hubbard model.
- $|\chi_i\rangle$: the i^{th} single-particle basis state ($i = 1, 2, \dots, N$). For example, $|\chi_{\mathbf{G}}\rangle$ can be a plane wave basis state with $\chi_{\mathbf{G}}(\mathbf{r}) \propto e^{i\mathbf{G}\cdot\mathbf{r}}$, where \mathbf{r} is a real-space coordinate.
- c_i^\dagger and c_i : creation and annihilation operators for an electron in $|\chi_i\rangle$. $n_i \equiv c_i^\dagger c_i$ is the corresponding number operator.
- M : number of electrons (if we omit spin index, e.g., if the system is fully polarized). In the more general case, M_σ is the number of electrons with spin σ ($\sigma = \uparrow$ or \downarrow). Of course, the choice of N above must ensure that $M_\sigma \leq N$.
- φ_m : single-particle orbital (we include an index m for discussions below to distinguish different single-particle orbitals). A single-particle orbital φ_m , given in terms of the single-particle basis states $\{|\chi_i\rangle\}$ as $\sum_i \varphi_{i,m} |\chi_i\rangle = \sum_i c_i^\dagger \varphi_{i,m} |0\rangle$, can be conveniently expressed as an N -dimensional vector:

$$\begin{pmatrix} \varphi_{1,m} \\ \varphi_{2,m} \\ \vdots \\ \varphi_{N,m} \end{pmatrix}$$

Given M different single-particle orbitals, we form a many-body wave function from their antisymmetrized product:

$$|\phi\rangle \equiv \hat{\varphi}_1^\dagger \hat{\varphi}_2^\dagger \cdots \hat{\varphi}_M^\dagger |0\rangle \quad (9)$$

where the operator

$$\hat{\varphi}_m^\dagger \equiv \sum_i c_i^\dagger \varphi_{i,m} \quad (10)$$

creates an electron in the m^{th} single-particle orbital $\{\varphi_{1,m}, \varphi_{2,m}, \dots, \varphi_{N,m}\}$. The many-body state $|\phi\rangle$ in Eq. (9) can be conveniently expressed as an $N \times M$ matrix:

$$\Phi \equiv \begin{pmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,M} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,M} \\ \vdots & \vdots & & \vdots \\ \varphi_{N,1} & \varphi_{N,2} & \cdots & \varphi_{N,M} \end{pmatrix}$$

Each column of this matrix represents a single-particle orbital that is completely specified by its N -dimensional vector. It is straightforward to keep the different columns orthonormalized. If the real-space coordinates of the electrons are $R = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M\}$, the many-body state in Eq. (9) gives (see Sec. 2.4)

$$\langle R|\phi\rangle = \phi(R) = \det \begin{pmatrix} \varphi_1(\mathbf{r}_1) & \varphi_2(\mathbf{r}_1) & \cdots & \varphi_M(\mathbf{r}_1) \\ \varphi_1(\mathbf{r}_2) & \varphi_2(\mathbf{r}_2) & \cdots & \varphi_M(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \varphi_1(\mathbf{r}_M) & \varphi_2(\mathbf{r}_M) & \cdots & \varphi_M(\mathbf{r}_M) \end{pmatrix},$$

where $\varphi_m(\mathbf{r}) = \sum_i \varphi_{i,m} \chi_i(\mathbf{r})$.

The many-body state $|\phi\rangle$ is known as a Slater determinant. One example of a Slater determinant is the HF solution $|\phi_{\text{HF}}\rangle = \prod_{\sigma} |\phi_{\text{HF}}^{\sigma}\rangle$, where $|\phi_{\text{HF}}^{\sigma}\rangle$ is defined by a matrix $\Phi_{\text{HF}}^{\sigma}$ whose columns are the M_{σ} lowest HF eigenstates. We can now add the following to our list of notations above:

- $|\phi\rangle$: a many-body wave function which can be written as a Slater determinant.
- Φ : an $N \times M$ matrix which represents the Slater determinant $|\phi\rangle$. Φ_{ij} will denote the matrix element of the matrix Φ in the i^{th} row and j^{th} column. For example, $\Phi_{ij} = \varphi_{i,j}$ above in Φ . Below when a Slater determinant $|\phi\rangle$ is referred to, it will often be helpful to think in terms of the matrix representation Φ operationally.
- $|\Psi\rangle$: a many-body wave function which is not necessarily a single Slater determinant, e.g., $|\Psi^{(n)}\rangle$ in Eq. (5).

Several properties of the Slater determinant are worth mentioning. For any two non-orthogonal Slater determinants, $|\phi\rangle$ and $|\phi'\rangle$, it can be shown that their overlap integral is

$$\langle \phi | \phi' \rangle = \det(\Phi^{\dagger} \Phi'). \quad (11)$$

The operation on any Slater determinant by any operator \hat{B} of the form

$$\hat{B} = \exp\left(\sum_{ij} c_i^{\dagger} U_{ij} c_j\right) \quad (12)$$

simply leads to another Slater determinant [34], i.e.,

$$\hat{B}|\phi\rangle = \hat{\phi}'_1 \hat{\phi}'_2 \cdots \hat{\phi}'_M |0\rangle \equiv |\phi'\rangle \quad (13)$$

with $\hat{\phi}'_m = \sum_j c_j^{\dagger} \Phi'_{jm}$ and $\Phi' \equiv e^U \Phi$, where U is a square matrix whose elements are given by U_{ij} and $B \equiv e^U$ is therefore an $N \times N$ square matrix as well. In other words, the operation of \hat{B} on $|\phi\rangle$ simply involves multiplying an $N \times N$ matrix by an $N \times M$ matrix.

We can define the expectation of an operator \hat{O} with respect to a pair of non-orthogonal Slater determinants:

$$\overline{\langle \hat{O} \rangle} \equiv \frac{\langle \phi | \hat{O} | \phi' \rangle}{\langle \phi | \phi' \rangle}. \quad (14)$$

The “bar” distinguishes it from the true interacting many-body expectations in Eq. (6), which we wish to compute with QMC. The simplest example of Eq. (14) is the single-particle Green’s function $G_{ij} \equiv \langle c_i c_j^\dagger \rangle$. In the ground-state formalism,

$$G_{ij} \equiv \frac{\langle \phi | c_i c_j^\dagger | \phi' \rangle}{\langle \phi | \phi' \rangle} = \delta_{ij} - [\Phi' (\Phi^\dagger \Phi')^{-1} \Phi^\dagger]_{ij}. \quad (15)$$

Given the Green’s function G , the general expectation defined in Eq. (14) can be computed for most operators of interest. This is an important property that will be used in our QMC calculations. For example, we can calculate the expectation of a general two-body operator, $\hat{O} = \sum_{ijkl} O_{ijkl} c_i^\dagger c_j^\dagger c_k c_l$, under the definition of Eq. (14):

$$\langle \hat{O} \rangle = \sum_{ijkl} O_{ijkl} (G'_{jk} G'_{il} - G'_{ik} G'_{jl}), \quad (16)$$

where the matrix G' is defined as $G' \equiv \mathbb{1} - G$.

Recently the use of a projected BCS wave function as a trial wave function in AFQMC has been formulated [35]. The overlap between a projected BCS wave function $|\Psi_{\text{BCS}}\rangle$ and a Slater determinant $|\phi\rangle$, and the corresponding “Green’s function” of Eq. (15) can be evaluated at a cost similar to that between two Slater determinants.

2.3 Hubbard-Stratonovich transformation

In order to carry out Eq. (5) in the Slater-determinant space we have introduced above, we write the many-body propagator $e^{-\Delta\tau \hat{H}}$ in single-particle form. With a small $\Delta\tau > 0$, the Trotter approximation can be used:

$$e^{-\Delta\tau \hat{H}} \approx e^{-\Delta\tau \hat{H}_1} e^{-\Delta\tau \hat{H}_2}, \quad (17)$$

which introduces a Trotter error. For actual implementation of the algorithms we discuss here, higher order Trotter break-ups are often used. The Trotter error can be further reduced with an extrapolation procedure after separate calculations have been done with different values of $\Delta\tau$. The \hat{H}_1 part of the propagator in (17) is the exponential of a one-body operator. The \hat{H}_2 part is not. It is, however, possible to rewrite $e^{-\Delta\tau \hat{H}_2}$ in this desired form through a so-called Hubbard-Stratonovich (HS) transformation [36]. For example, for the Hubbard model in Eq. (3), an exact, discrete HS transformation [37] exists for the repulsive on-site repulsion

$$e^{-\Delta\tau U n_{i\uparrow} n_{i\downarrow}} = e^{-\Delta\tau U (n_{i\uparrow} + n_{i\downarrow})/2} \sum_{x_i=\pm 1} \frac{1}{2} e^{\gamma x_i (n_{i\uparrow} - n_{i\downarrow})}, \quad (18)$$

where the constant γ is determined by $\cosh(\gamma) = \exp(\Delta\tau U/2)$. Similarly, for an attractive interaction $V n_{i\uparrow} n_{i\downarrow}$ with $V < 0$

$$e^{-\Delta\tau V n_{i\uparrow} n_{i\downarrow}} = e^{-\Delta\tau V (n_{i\uparrow} + n_{i\downarrow} - 1)/2} \sum_{x_i=\pm 1} \frac{1}{2} e^{\gamma x_i (n_{i\uparrow} + n_{i\downarrow} - 1)}, \quad (19)$$

where $\cosh(\gamma) = \exp(\Delta\tau |V|/2)$.

In general, if \hat{H}_2 can be written as a sum of squares of one-body operators, of the form

$$\hat{H}_2 = \frac{1}{2} \sum_{\gamma=1}^{N_\gamma} \lambda_\gamma \hat{v}_\gamma^2, \quad (20)$$

we can apply the HS transformation to each term (after Trotter break-up):

$$e^{-(\Delta\tau/2) \lambda \hat{v}^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-x^2/2} e^{x\sqrt{-\Delta\tau\lambda} \hat{v}}, \quad (21)$$

where x is an auxiliary-field variable. The constant in front of \hat{v} in the exponent on the right-hand side can be real or imaginary, depending on the sign of λ . The key is that the quadratic form (in \hat{v}) on the left is replaced by a linear one on the right. In Eq. (20), λ_γ is a constant and \hat{v}_γ is a one-body operator similar to \hat{H}_1 . Each γ leads to one auxiliary-field in Eq. (21).

We now show one way to manipulate a general \hat{H}_2 into the form in Eq. (20). Let us cast V_{ijkl} in Eq. (2) in the form of a Hermitian matrix by introducing two indices $\alpha = (i, l)$ and $\beta = (k, j)$: $\mathcal{V}_{\alpha\beta} = \mathcal{V}_{(i,l),(k,j)} = V_{ijkl}$. The Hermitian matrix \mathcal{V} can then be diagonalized and written as $\mathcal{V} = R\Lambda R^\dagger$, where R is a matrix whose columns are the eigenvectors of \mathcal{V} and Λ is a diagonal matrix containing the corresponding eigenvalues λ_α . That is

$$\mathcal{V}_{\alpha\beta} = \sum_{\gamma} R_{\alpha\gamma} \lambda_\gamma R_{\beta\gamma}^*. \quad (22)$$

The two-body operator \hat{V} can therefore be written as

$$\begin{aligned} \hat{H}_2 &= \sum_{ijkl} V_{ijkl} c_i^\dagger c_l c_j^\dagger c_k - \sum_{ijkl} V_{ijkl} c_i^\dagger c_k \delta_{jl} \\ &= \sum_{\gamma} \lambda_\gamma \left(\sum_{il} R_{(i,l)\gamma} c_i^\dagger c_l \right) \left(\sum_{jk} R_{(k,j)\gamma}^* c_j^\dagger c_k \right) - \sum_{ik} \left(\sum_j V_{ijkj} \right) c_i^\dagger c_k. \end{aligned}$$

Noting that \hat{H}_2 is Hermitian, we can put the above in a more symmetric form

$$\hat{H}_2 = \frac{1}{2} \sum_{\gamma} \lambda_\gamma \{ \hat{\rho}_\gamma, \hat{\rho}_\gamma^\dagger \} + \hat{\rho}_0, \quad (23)$$

where the one-body operators are defined as $\hat{\rho}_\gamma \equiv \sum_{il} R_{(i,l)\gamma} c_i^\dagger c_l$ and $\hat{\rho}_0 \equiv -\sum_{ik} [\sum_j (V_{ijkj} + V_{jikj})/2] c_i^\dagger c_k$. Since

$$\{ \hat{\rho}_\gamma, \hat{\rho}_\gamma^\dagger \} = \frac{1}{2} [(\hat{\rho}_\gamma + \hat{\rho}_\gamma^\dagger)^2 - (\hat{\rho}_\gamma - \hat{\rho}_\gamma^\dagger)^2] \equiv \hat{v}_\gamma^2 - \hat{v}_\gamma'^2, \quad (24)$$

we have succeeded in writing \hat{H}_2 in the desired form.

The way to decompose \hat{H}_2 above leads to approximately $2N^2$ auxiliary fields. Often the interaction simplifies \hat{H}_2 and the number of auxiliary fields can be much reduced. For example, straightforward applications to molecular systems using Gaussian basis sets employed the above decomposition, except that the matrix elements are all real [21, 38] (and terms with λ below a

small threshold can be discarded). A modified Cholesky decomposition has been used which leads to $\mathcal{O}(N)$ fields [32]. In a plane-wave basis, the two-body part is the Fourier transform of $1/|\mathbf{r}_m - \mathbf{r}_n|$ which leads to $1/\mathbf{Q}^2$ [27].

Different forms of the HS transformation exist [39], and can affect the performance of the QMC method. For example, subtracting a mean-field “background” from the two-body term prior to the decomposition improves the calculations [21, 31, 40]. Indeed experience shows that they can not only impact the statistical accuracy, but also lead to different quality of approximations under the constrained-path methods that we discuss below. Therefore, although we discuss the algorithms under the generic form of Eq. (25), we emphasize that it is worthwhile to explore different forms of the HS transformation in an actual implementation [41].

If we denote the collection of auxiliary fields by \mathbf{x} and combine one-body terms from \hat{H}_1 and \hat{H}_2 , we obtain the following compact representation of the outcome of the HS transformation:

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x} p(\mathbf{x}) \hat{B}(\mathbf{x}), \quad (25)$$

where $p(\mathbf{x})$ is a probability density function (PDF), for example, a multi-dimensional Gaussian. The propagator $\hat{B}(\mathbf{x})$ in Eq. (25) has a *special form*, namely, it is a product of operators of the type in Eq. (12), with U_{ij} depending on the auxiliary field. The matrix representation of $\hat{B}(\mathbf{x})$ will be denoted by $B(\mathbf{x})$.

In essence, the HS transformation replaces the two-body interaction by one-body interactions with a set of random external auxiliary fields. In other words, it converts an interacting system into many *non-interacting* systems living in fluctuating external auxiliary-fields. The sum over all configurations of auxiliary fields recovers the interaction.

2.4 A simple example

We now study in some details the ground-state of the one-dimensional Hubbard model with 4 sites and with open boundary conditions. We can think of the lattice as a crude representation of particles in a one-dimensional box, with only four equally-spaced grid points to discretize the continuous space inside the box and the kinetic energy written in terms of finite difference. We will consider $N_\uparrow = 2$ and $N_\downarrow = 1$. This can be thought of as a chain of 4 hydrogen atoms, with one of them ionized to H^+ , in a minimal basis. We label the four sites 1 through 4.

First let us examine the trivial case of free electrons, i.e. $U = 0$. We can write down the *1-electron* Hamiltonian matrix, based on Eq. (3), which is of dimension 4×4 :

$$H = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \quad (26)$$

The eigenstates of H can be obtained by direct diagonalization. With these eigenstates, we immediately obtain the ground-state wave function $|\psi_0\rangle$ of the 3-electron system from the Pauli

exclusion principle:

$$|\psi_0\rangle = \begin{pmatrix} 0.3717 & -0.6015 \\ 0.6015 & -0.3717 \\ 0.6015 & 0.3717 \\ 0.3717 & 0.6015 \end{pmatrix} \otimes \begin{pmatrix} 0.3717 \\ 0.6015 \\ 0.6015 \\ 0.3717 \end{pmatrix},$$

where the first matrix contains two single-particle orbitals (two columns) for the two \uparrow electrons and the second matrix contains one single-electron orbital for the one \downarrow electron. Each single-electron orbital is an eigenvector of H . Note the second and third lowest single-electron states are degenerate, causing the 3-electron system to be open shell and $|\psi_0\rangle$ to be degenerate. We have simply picked one particular linear combination for the second \uparrow -electron in $|\psi_0\rangle$ above.

An object of the form of $|\psi_0\rangle$ is of course nothing more than a Slater determinant. For example, the amplitude of the configuration $|R\rangle = |\downarrow \uparrow 0 \uparrow\rangle$, i.e., two \uparrow electrons on sites 2 and 4 and the one \downarrow electron on site 1, is given by

$$\langle R|\psi_0\rangle = \det \begin{pmatrix} 0.6015 & -0.3717 \\ 0.3717 & 0.6015 \end{pmatrix} \cdot \det \begin{pmatrix} 0.3717 \end{pmatrix}.$$

That is, more formally,

$$|R\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and $\langle R|\psi_0\rangle = \det(R^\dagger \cdot \Psi_0)$ is a number, where R and Ψ_0 denote the matrices corresponding to $|R\rangle$ and $|\psi_0\rangle$, respectively.

For this non-interacting system, an alternative (albeit indirect and indeed circular) way of obtaining $|\psi_0\rangle$ is by the power method. From the eigenvalues and eigenvectors of \hat{K} , we can easily construct the matrix for $e^{-\Delta\tau\hat{K}}$, which has the same structure as H in Eq. (26) above (i.e., a 4×4 matrix $e^{-\Delta\tau H}$). Denote this matrix by B_K . With an arbitrarily chosen initial Slater determinant $|\psi^{(0)}\rangle$ (with non-zero overlap with $|\psi_0\rangle$), we can then repeatedly apply $e^{-\Delta\tau\hat{H}}$ to carry out the iterative process in Eq. (5), which means multiplying both the 4×2 \uparrow matrix and the 4×1 \downarrow matrix by B_K . The process will lead to $|\psi_0\rangle$ as $n \rightarrow \infty$.

Note that the solution remains a single Slater determinant during the iteration, as discussed at the end of Sec. 2.1. We could obtain the self-consistent solution of a mean-field equation via this procedure, which looks like a (imaginary-)time-dependent HF or DFT calculation. The orbitals in the determinant (in this example the two \uparrow -spin orbitals) must be periodically re-orthonormalized to avoid the loss of numerical precision. This procedure is often done with a modified Gram-Schmidt, and is standard in mean-field calculations. Incidentally, the tendency for the orbitals to collapse to the lowest energy single-particle orbital is intimately related to the sign problem. In any simulation without the use of Slater determinant (second quantization) as its random walker, for example in standard DMC, this is the driving force for the sign problem.

Now let us consider the many-body case, by turning on the interaction U . The first approach of directly diagonalizing H is the method of exact diagonalization. The size of the matrix is now the size of the Hilbert space (not just N), which in this example is 24. Thus the computational cost of this approach grows exponentially with system size. The power method of Eq. (5), on the other hand, can still apply if we can write $e^{-\Delta\tau\hat{H}}$ in some one-electron form. The HS transformation does just that. Assuming $\Delta\tau$ is small and applying the Trotter break-up, we have

$$e^{-\Delta\tau\hat{H}} \Rightarrow \sum_{\mathbf{x}} p(\mathbf{x}) \begin{pmatrix} e^{\gamma x_1} & 0 & 0 & 0 \\ 0 & e^{\gamma x_2} & 0 & 0 \\ 0 & 0 & e^{\gamma x_3} & 0 \\ 0 & 0 & 0 & e^{\gamma x_4} \end{pmatrix} \cdot B_K \otimes \begin{pmatrix} e^{-\gamma x_1} & 0 & 0 & 0 \\ 0 & e^{-\gamma x_2} & 0 & 0 \\ 0 & 0 & e^{-\gamma x_3} & 0 \\ 0 & 0 & 0 & e^{-\gamma x_4} \end{pmatrix} \cdot B_K,$$

where $\mathbf{x} = \{x_1, x_2, x_3, x_4\}$. This is just Eq. (25). Note that $B(\mathbf{x})$ has an \uparrow and a \downarrow component, each of which is a 4×4 matrix. Applying each $B(\mathbf{x})$ to a Slater determinant means precisely the same as in the non-interacting case (with $B_K \otimes B_K$). In other words, $B(\mathbf{x})$ operating on any Slater determinant $|\phi\rangle$ simply involves matrix multiplications for the \uparrow and \downarrow components separately, leading to another Slater determinant $|\phi'\rangle$, as in Eq. (13). Starting from a single Slater determinant, e.g., $|\Psi^{(0)}\rangle = |\psi_0\rangle$, we would end up with a multi-determinant representation of $|\Psi^{(1)}\rangle$, and so on. We will maintain a constant population as described below.

3 Ground-State AFQMC Methods

In this section, we discuss the ground-state AFQMC method. The CPMC and its generalization to the phase problem are free of the decay of the average sign. The methods are approximate, relying on what we will generally refer to as the constrained path approximation.

3.1 Free-projection AFQMC

In this section we briefly describe the ground-state AFQMC method without any constraints. We will rely on the machinery established in the previous section. Our goal is to illustrate the essential ideas, in a way which will facilitate our discussion of the sign problem and help introduce the framework for the constrained path Monte Carlo methods. We will not go into details such as how to efficiently sample the auxiliary fields or how to stabilize the matrices. They are described in the literature. We write the usual path-integral and Metropolis form explicitly here to show that it is the same as the open-ended random walk approach we take in CPMC, as far as understanding the sign problem is concerned.

We want to compute ground-state expectation values $\langle\hat{O}\rangle$ using (6) and (25). The denominator in (6) is

$$\begin{aligned} \langle\psi^{(0)}|e^{-n\Delta\tau\hat{H}}e^{-n\Delta\tau\hat{H}}|\psi^{(0)}\rangle &= \int \langle\psi^{(0)}|\left[\prod_{l=1}^{2n} d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)}) \hat{B}(\mathbf{x}^{(l)})\right]|\psi^{(0)}\rangle \\ &= \int \left[\prod_l d\mathbf{x}^{(l)} p(\mathbf{x}^{(l)})\right] \det\left([\Psi^{(0)}]^\dagger \prod_l B(\mathbf{x}^{(l)}) \Psi^{(0)}\right). \end{aligned} \quad (27)$$

In the standard ground-state AFQMC method [13], a value of n is first chosen and fixed throughout the calculation. If we use X to denote the collection of the auxiliary-fields $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(2n)}\}$ and $D(X)$ to represent the integrand in Eq. (27), we can write the expectation value of Eq. (6) as

$$\langle \hat{O} \rangle = \frac{\int \overline{\langle \hat{O} \rangle} D(X) dX}{\int D(X) dX} = \frac{\int \overline{\langle \hat{O} \rangle} |D(X)| s(X) dX}{\int |D(X)| s(X) dX}, \quad (28)$$

where

$$s(X) \equiv D(X)/|D(X)| \quad (29)$$

measures the “sign” of $D(X)$. The non-interacting expectation value for a given X is that defined in Eq. (14)

$$\overline{\langle \hat{O} \rangle} \equiv \frac{\langle \phi_L | \hat{O} | \phi_R \rangle}{\langle \phi_L | \phi_R \rangle} \quad (30)$$

with

$$\begin{aligned} \langle \phi_L | &= \langle \psi^{(0)} | \hat{B}(\mathbf{x}^{(2n)}) \hat{B}(\mathbf{x}^{(2n-1)}) \dots \hat{B}(\mathbf{x}^{(n+1)}) \\ | \phi_R \rangle &= \hat{B}(\mathbf{x}^{(n)}) \hat{B}(\mathbf{x}^{(n-1)}) \dots \hat{B}(\mathbf{x}^{(1)}) | \psi^{(0)} \rangle, \end{aligned}$$

which are both Slater determinants.

$D(X)$ as well as $\langle \phi_L |$ and $| \phi_R \rangle$ are completely determined by the path X in auxiliary-field space. The expectation in Eq. (28) is therefore in the form of Eq. (62), with $f(X) = |D(X)|$ and $g(X) = \overline{\langle \hat{O} \rangle}$. The important point is that, for each X , $|D(X)|$ is a number and $g(X)$ can be evaluated using Eqn. (15) and (16). Often the Metropolis Monte Carlo algorithm [42] is used to sample auxiliary-fields X from $|D(X)|$. Any $\langle \hat{O} \rangle$ can then be computed following the procedure described by Eq. (61) in the appendix.

We will carry out free-projection calculations with an open-ended random walk [20, 40] similar to the CP approach, instead of using Metropolis sampling as outlined above [14, 43] which keeps an entire path of a *fixed length* $2n$ as the object to sample. (An example of free-projection is shown in Fig. 2 in Sec. 3.3.) The open-ended random walk framework does not have any ergodicity issues, and it is straightforward to project to longer imaginary-time in order to approach the ground state. This approach also allows a natural implementation of importance sampling and the constraint, as we discuss below.

3.2 Why and how does the sign problem occur?

The sign problem occurs because of the fundamental symmetry between the fermion ground-state $|\Psi_0\rangle$ and its negative $-|\Psi_0\rangle$ [19, 44]. For any ensemble of Slater determinants $\{|\phi\rangle\}$ which gives a Monte Carlo representation of the ground-state wave function, this symmetry implies that there exists another ensemble $\{-|\phi\rangle\}$ which is also a correct representation. In other words, the Slater determinant space can be divided into two degenerate halves (+ and -) whose bounding surface \mathcal{N} is defined by $\langle \Psi_0 | \phi \rangle = 0$. This surface is in general *unknown*.

In some special cases symmetry prohibits any crossing of \mathcal{N} in the random walk. The AFQMC calculation is then free of the sign problem. An example [45] is the half-filled ($M_\uparrow + M_\downarrow = N$) repulsive Hubbard model, which via particle-hole symmetry can be mapped to an attractive model with no spin polarization: $M_\uparrow = M_\downarrow$. It is easy to see this in the latter case: if we choose an initial population of walkers of the form $|\phi_\uparrow\rangle \otimes |\phi_\downarrow\rangle$ with $|\phi_\uparrow\rangle = |\phi_\downarrow\rangle$, the structure will be preserved by the propagation. (In general when there is a twist to the boundary condition, the symmetry between \uparrow - and \downarrow -spins should be of opposite momenta [35].) The overlap of each walker with a wave function $|\Psi_0\rangle$ which observes the same symmetry will be a square and thus always non-negative. In more general cases, however, walkers do cross \mathcal{N} in their propagation by $e^{-\Delta\tau\hat{H}}$. The sign problem then invariably occurs.

In Fig. 1, we illustrate the space of Slater determinants by a one-dimensional (horizontal) line. The “node” \mathcal{N} is the (red) dot in the middle. Imaginary time (or n) is in the vertical direction and increases as the arrow suggests. That is, as the walker moves in the horizontal line, we stretch out continuously “snapshots” of its position along the vertical direction. Now we follow an initial Slater determinant. With no loss of generality, we assume it has a positive overlap with $|\Psi_0\rangle$. At time $n = 0$ it is indicated by the (green) dot on the right. As the random walk evolves, the walker can reach the node, which is the (red) vertical line. At the instant it lands on \mathcal{N} , the walker will make no further contribution to the representation of the ground state, since

$$\langle\Psi_0|\phi\rangle = 0 \Rightarrow \langle\Psi_0|e^{-\beta\hat{H}}|\phi\rangle = 0 \text{ for any } \beta. \quad (31)$$

Paths that result from such a walker have equal probability of being in either half of the Slater determinant space. A few of these possible paths are shown by dashed lines. Computed analytically, they would cancel and not make any contribution in the ground-state wave function, as indicated by their symmetric placement with respect to the node line. But since the random walk has no knowledge of \mathcal{N} , these paths continue to be sampled (randomly) in the random walk and become MC noise. Only paths that are completely confined to the right-hand side, as shown by the solid (green) line, will lead to contributions to the ground state, but the relative number of such confined paths decreases exponentially with n . Asymptotically in n , the MC representation of the ground-state wave function consists of an *equal* mixture of the $+$ and $-$ walkers, regardless of where the random walks originated. The Monte Carlo signal is therefore lost. The decay of the signal-to-noise ratio, i.e. the decay of the average sign of $\langle\Psi_T|\phi\rangle$, occurs at an exponential rate with imaginary time.

To eliminate the decay of the signal-to-noise ratio, we impose the constrained-path approximation. Fahy and Hamann first used [25] such a constraint in the framework of the standard AFQMC method. However, the non-local nature of such a constraint proved difficult to implement efficiently in the “path-integral-like” scheme. Here, with the open-ended random walk formalism, the constraint only needs to be imposed one time-step at a time and is extremely simple to implement. It requires that each random walker at each step have a positive overlap with the trial wave function $|\Psi_T\rangle$:

$$\langle\Psi_T|\phi_k^{(n)}\rangle > 0. \quad (32)$$

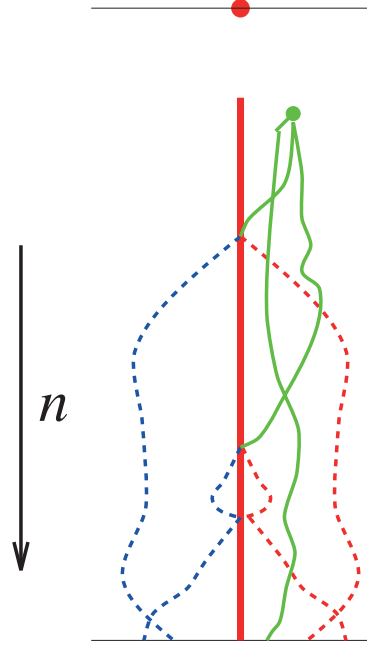


Fig. 1: Schematic illustration of the sign problem [19]. The top line represents Slater determinant space; the dot represents the “node” \mathcal{N} , where a determinant is orthogonal to the ground state $|\Psi_0\rangle$. As the projection continues (increasing n), Slater determinants undergo random walks, tracing out “paths” as shown. When a walker reaches \mathcal{N} , its future paths will collectively cancel in their contribution to $|\Psi_0\rangle$, indicated by the symmetric distribution of dashed paths about the nodal line. The Monte Carlo sampling, with no knowledge of this cancellation, continues to sample such paths randomly. The relative number of paths with real contributions (solid paths) decreases exponentially as n increases.

This yields an approximate solution to the ground-state wave function, $|\Psi_0^c\rangle = \sum_{\phi} |\phi\rangle$, in which all Slater determinants $|\phi\rangle$ satisfy (32).

From (31), it follows that the constrained path approximation becomes *exact* for an exact trial wave function $|\Psi_T\rangle = |\Psi_0\rangle$. The overall normalization of walkers remains a constant on the average, since the loss of walkers at \mathcal{N} is compensated by the branching of walkers elsewhere; that is, the eigenvalue problem with \mathcal{N} as boundary has a stable solution.

To implement the constrained-path approximation in the random walk, we define the importance function (see next section) by:

$$O_T(\phi) \equiv \max\{\langle\Psi_T|\phi\rangle, 0\}. \quad (33)$$

This prevents walkers from crossing the trial nodal surface \mathcal{N} and entering the “−” half-space as defined by $|\Psi_T\rangle$. A walker k in the n^{th} -step is eliminated (assigned weight zero) for the $(n+1)^{\text{st}}$ -step when its $O_T(\phi_k^{(n+1)})$ becomes zero. In the limit $\Delta\tau \rightarrow 0$, Eq. (33) ensures that the walker distribution vanishes smoothly at \mathcal{N} and the constrained-path approximation is properly imposed. With a finite $\Delta\tau$, the procedure eliminates walker k when $O_T(\phi_k^{(n+1)}) \leq 0$. The error that results is a form of Trotter error [24] which can be further reduced using what is known as a mirror correction [20, 24].

3.3 The constrained path Monte Carlo method

The constrained path Monte Carlo (CPMC) method [19, 20, 46] incorporates the idea discussed above into a natural and efficient algorithm. The approach applies to Hubbard-like Hamiltonians where the auxiliary-field is real. Even in the case where a twist angle is applied to the boundary condition and the hopping matrix elements are complex, a simple generalization of this condition is sufficient [46]. Using (25), we write Eq. (5) as

$$|\Psi^{(n+1)}\rangle = \int d\mathbf{x} p(\mathbf{x}) \hat{B}(\mathbf{x}) |\Psi^{(n)}\rangle. \quad (34)$$

In the random walk realization of this iteration, we represent the wave function at each stage by a finite ensemble of Slater determinants, i.e.,

$$|\Psi^{(n)}\rangle \propto \sum_k w_k^{(n)} |\phi_k^{(n)}\rangle, \quad (35)$$

where k labels the Slater determinants and an overall normalization factor of the wave function has been omitted. A weight factor $w_k^{(n)}$ is introduced for each walker, even though in Eq. (34) the kernel p is normalized. This is because single-particle orbitals in a Slater determinants cease to be orthogonal to each other as a result of propagation by \hat{B} . When they are re-orthogonalized [19, 20], an overall factor appears, which we will view as the w term in the integral equation Eq. (64) of the appendix.

The structure of the random walk now resembles that of Eq. (64). For each random walker we sample an auxiliary-field configuration \mathbf{x} from the PDF $p(\mathbf{x})$ and propagate the walker to a new one via $\Phi_k^{(n+1)} = B(\mathbf{x})\Phi_k^{(n)}$. If necessary, a re-orthogonalization procedure is applied to $\Phi_k^{(n)}$ prior to the propagation: $\Phi_k^{(n)} = [\Phi_k^{(n)}]'R$, where R is an $M \times M$ upper-triangular matrix. $[\Phi_k^{(n)}]'$ is then used in the propagation instead; the weight of the new walker is $w_k^{(n+1)} = w_k^{(n)} \det(R)$. The simple random walk procedure is sufficient and convenient for thinking about many conceptual issues. It is, however, not efficient as a practical algorithm in most cases of interest, because the sampling of \mathbf{x} is completely random with no regard to the potential contribution to $D(X)$. The idea of importance sampling is to iterate a modified equation with a modified wave function, without changing the underlying eigenvalue problem of (34). Specifically, for each Slater determinant $|\phi\rangle$, we define an importance function as given in Eq. (33) that estimates its overlap with the ground-state wave function. We can rewrite Eq. (34) as

$$|\tilde{\Psi}^{(n+1)}\rangle = \int d\mathbf{x} \tilde{p}(\mathbf{x}) \hat{B}(\mathbf{x}) |\tilde{\Psi}^{(n)}\rangle, \quad (36)$$

where the modified ‘‘PDF’’ is

$$\tilde{p}(\mathbf{x}) = \frac{O_T(\phi^{(n+1)})}{O_T(\phi^{(n)})} p(\mathbf{x}). \quad (37)$$

With the new kernel \tilde{p} , the probability distribution for \mathbf{x} vanishes smoothly as $O_T(\hat{B}(\mathbf{x})\phi^{(n)})$ approaches zero, and the constraint is naturally imposed. As expected, $\tilde{p}(\mathbf{x})$ is a function of both the current and the next position in Slater-determinant space. Further, it modifies $p(\mathbf{x})$

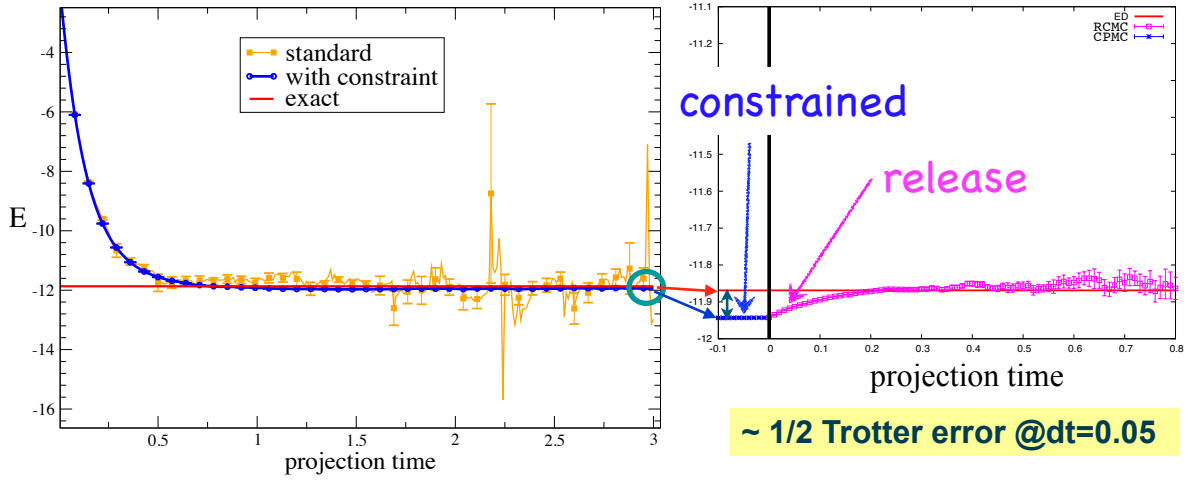


Fig. 2: Controlling the sign problem with a systematically improvable approximation. The total energy is plotted vs. imaginary-time for a 4×4 Hubbard model with $N_\uparrow = N_\downarrow = 7$ electrons (i.e., density $n = 0.875$) and $U/t = 8$, which is in the regime important for magnetism and superconductivity, and where the sign problem is severe. The **left panel** shows the sign problem in free-projection [15, 22] (yellow), results with constraint [20, 46] (blue), and exact diagonalization (magenta). The **right panel** illustrates constraint release [41]. To the left of the vertical bar, a magnified view is shown of the exact and constrained path results. The error from the constraint is roughly half of the Trotter error with the typically recommended value of $\Delta\tau = 0.05$. Releasing the constraint removes the small bias and converges to the exact answer, with growing computational cost, as seen on the right.

such that the probability is increased when \mathbf{x} leads to a determinant with larger overlap and is decreased otherwise. It is trivially verified that Eqs. (34) and (36) are identical. If $|\Psi_T\rangle = |\Psi_0\rangle$, the normalization $\int \tilde{p}(\mathbf{x}) d\mathbf{x}$ becomes a constant, which means that the weights of walkers will remain a constant and the random walk will have no fluctuations, as should be the case with perfect importance sampling.

In the random walk, the ensemble of walkers $\{|\phi_k^{(n)}\rangle\}$ now represents the modified wave function: $|\tilde{\Psi}^{(n)}\rangle \propto \sum_k w_k^{(n)} |\phi_k^{(n)}\rangle$. The true wave function is then given formally by

$$|\Psi^{(n)}\rangle \propto \sum_k w_k^{(n)} \frac{|\phi_k^{(n)}\rangle}{O_T(\phi_k^{(n)})}, \quad (38)$$

although in actual measurements it is $|\tilde{\Psi}^{(n)}\rangle$ that is needed and division by O_T does not appear. The random walk process is similar to that discussed for Eq. (34), but with $p(\mathbf{x})$ replaced by $\tilde{p}(\mathbf{x})$. The latter is in general not a normalized PDF, and we denote the normalization constant for walker k by $N(\phi_k^{(n)})$ and rewrite the iterative relation as

$$|\phi_k^{(n+1)}\rangle \leftarrow N(\phi_k^{(n)}) \int d\mathbf{x} \frac{\tilde{p}(\mathbf{x})}{N(\phi_k^{(n)})} B(\mathbf{x}) |\phi_k^{(n)}\rangle. \quad (39)$$

Thus, for each walker $|\phi_k^{(n)}\rangle$, one step of the random walk consists of:

1. sampling an \mathbf{x} from the PDF $\tilde{p}(\mathbf{x})/N(\phi_k^{(n)})$. With discrete, Ising-like auxiliary-fields for a Hubbard interaction, the sampling is achieved by a heatbath-like algorithm, sweeping through each field x_i [20]
2. constructing the corresponding $B(\mathbf{x})$ and then propagating the walker $\Phi_k^{(n)}$ to generate a new walker
3. assigning a weight $w_k^{(n+1)} = w_k^{(n)} N(\phi_k^{(n)})$ to the new walker.

In contrast with the primitive algorithm in Eq. (34), the weight of a walker does not need to be modified here when the re-orthogonalization procedure is applied. This is because the upper-triangular matrix R only contributes to the overlap O_T , which is already represented by the walker weight. After each re-orthonormalization [24, 47], R can simply be discarded [20, 24]. To calculate the expectation value of an observable which does not commute with the Hamiltonian, the back-propagation technique [20, 30] is used. The idea is to create two coupled populations to represent the bra and ket in Eq. (6), respectively. Because the population in the random walk is importance-sampled, two independent populations which are uncoupled would lead to large fluctuations in the estimator after the importance functions have been "undone." In back-propagation, we choose an iteration n and store the entire population $\{|\phi_k^{(n)}\rangle\}$. As the random walk proceeds from n , we keep track of the following two items for each new walker: (1) the sampled auxiliary-field values that led to the new walker from its parent walker and (2) an integer label that identifies the parent. After an additional m iterations, we carry out the back-propagation: For each walker l in the $(n + m)^{\text{th}}$ (current) population, we initiate a determinant $\langle\psi_T|$ and act on it with the corresponding propagators, but taken in reverse order. The m successive propagators are constructed from the items stored between steps m and $m + l$, with $\exp(-\Delta\tau\hat{H}_1/2)$ inserted where necessary. The resulting determinants $\langle\bar{\phi}_l^{(m)}|$ are combined with its parent from iteration n to compute $\langle\mathcal{O}\rangle_{\text{BP}}$, in a way similar to the mixed estimator. The weights are given correctly by $w_l^{(n+m)}$, due to importance sampling [30]. The weight works similarly to that in forward walking in DMC. However, the trial wave function can be propagated backwards in AFQMC to obtain a population sampling the bra, from which the entire single-particle Green's function can be computed with Eq. (15). Thus AFQMC can compute both diagonal and off-diagonal observables.

3.4 The phaseless formalism for complex auxiliary-fields

When the many-body Hamiltonian leads to a decomposition with $\lambda > 0$ in Eq. (20), the resulting HS transformation will have complex auxiliary-fields. This is the case for the electron-electron repulsion. (We will sometimes refer loosely to having complex auxiliary-fields. It should be understood that this means that the propagator resulting from the two-body Hamiltonian is complex. Incidentally, it is always possible to have real auxiliary-fields, for example by making a negative shift to the positive potential, but that simply leads to many fluctuating fields to recover a constant background, and a much more severe sign problem [15, 41].) In this situation

a phase problem arises, as illustrated in Fig. 3. We next describe the formalism to deal with the problem.

With a continuous auxiliary-field, the importance sampling in step 1 in Sec. 3.3 is achieved with a force bias [15, 30]. To sketch a derivation we write the two-body propagator as

$$\int e^{-\mathbf{x}^2/2} e^{\mathbf{x} \cdot \hat{\mathbf{v}}} d\mathbf{x}, \quad (40)$$

where $\hat{\mathbf{v}}$ is a vector denoting the collection of one-body operators labeled by γ in Eq. (20), and the constant $\sqrt{-\Delta\tau\lambda_\gamma}$, which is complex if λ_γ is positive from in Eq. (21), has been absorbed into the definition of $\hat{\mathbf{v}}$. We introduce a shift in the integral in Eq. (40), which leads to an alternative propagator

$$\int e^{-\mathbf{x}^2/2} e^{\mathbf{x} \cdot \bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} e^{(\mathbf{x} - \bar{\mathbf{x}}) \cdot \hat{\mathbf{v}}} d\mathbf{x}. \quad (41)$$

The new propagator is exact for any choice of the shift $\bar{\mathbf{x}}$, which can be complex in general.

In free projection, application of the propagator in Eq. (41) to a Slater determinant $|\phi\rangle$ would be similar to that discussed in the previous section. We would sample \mathbf{x} from the Gaussian, propagate $|\phi\rangle$ to a new $|\phi'(\mathbf{x})\rangle$ by

$$|\phi'(\mathbf{x})\rangle = e^{(\mathbf{x} - \bar{\mathbf{x}}) \cdot \hat{\mathbf{v}}} |\phi\rangle, \quad (42)$$

and include the extra exponential with $\bar{\mathbf{x}}$ in the weight of the walker.

In the iterative projection, the ground-state wave function that emerges after the equilibration phase can be written in the following form

$$|\Psi_0\rangle = \int f(\phi) |\phi\rangle d\phi, \quad (43)$$

where we assume a criterion has been chosen and is followed consistently to separate the determinant $|\phi\rangle$ from its coefficient (e.g., orthonormalization). Because of the over-completeness of the basis space, the function $f(\phi)$ may be different in different steps of the iteration, but in each step it is uniquely determined. In free projection the function $f(\phi)$ is represented by the weights of the MC samples, and the MC representation of the ground-state wave function is given by Eq. (35). Omitting the explicit walker index, we have

$$|\Psi_0\rangle = \sum_{\phi} w_{\phi} |\phi\rangle, \quad (44)$$

where the sum runs over all walkers and w_{ϕ} is the weight of walker $|\phi\rangle$. In general, the weight w_{ϕ} can be complex (as can the orbitals), which is of course the cause for the phase problem.

Using the idea of importance sampling, we further modify the propagator in Eq. (41)

$$\int \langle \Psi_T | \phi'(\mathbf{x}) \rangle e^{-\mathbf{x}^2/2} e^{\mathbf{x} \cdot \bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} e^{(\mathbf{x} - \bar{\mathbf{x}}) \cdot \hat{\mathbf{v}}} \frac{1}{\langle \Psi_T | \phi \rangle} d\mathbf{x}, \quad (45)$$

where, as before, the trial wave function $|\Psi_T\rangle$ represents the best guess to $|\Psi_0\rangle$ and is the one used in the mixed estimate for observables. The new propagator means that the walker weights

now represent, instead of $f(\phi)$ in Eq. (43), a modified function $f(\phi)\langle\Psi_T|\phi\rangle$. In contrast with Eq. (44), the MC representation of the ground-state wave function with importance sampling is now given by Eq. (38)

$$|\Psi_0\rangle = \sum_{\phi} w_{\phi} \frac{|\phi\rangle}{\langle\Psi_T|\phi\rangle}, \quad (46)$$

where any overall phase of the walker is removed.

We now consider the ratio $\langle\Psi_T|\phi'(\mathbf{x})\rangle/\langle\Psi_T|\phi\rangle$ in Eq. (45). Defining

$$\bar{\mathbf{v}} \equiv -\frac{\langle\Psi_T|\hat{\mathbf{v}}|\phi\rangle}{\langle\Psi_T|\phi\rangle} \sim \mathcal{O}(\sqrt{\Delta\tau}) \quad (47)$$

and

$$\overline{\mathbf{v}^2} \equiv \frac{\langle\Psi_T|\hat{\mathbf{v}}^2|\phi\rangle}{\langle\Psi_T|\phi\rangle} \sim \mathcal{O}(\Delta\tau), \quad (48)$$

we can evaluate the ratio by expanding the propagator [48] in Eq. (42) to $\mathcal{O}(\tau)$. The overall weight factor in Eq. (45) can now be written in the form of an exponential [15, 30]

$$\frac{\langle\Psi_T|\phi'(\mathbf{x})\rangle}{\langle\Psi_T|\phi\rangle} e^{\mathbf{x}\cdot\bar{\mathbf{x}}-\bar{\mathbf{x}}^2/2} \doteq \exp[-(\mathbf{x}-\bar{\mathbf{x}})\cdot\bar{\mathbf{v}} + \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\overline{\mathbf{v}^2} - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^2\bar{\mathbf{v}}^2 + \mathbf{x}\cdot\bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2]. \quad (49)$$

The optimal choice of the shift $\bar{\mathbf{x}}$, which minimizes the fluctuation of Eq. (49) with respect to \mathbf{x} , is therefore $\bar{\mathbf{x}} = \bar{\mathbf{v}}$. With this choice, the weight factor in Eq. (49) is $\exp[\mathbf{x}^2\overline{\mathbf{v}^2}/2 + (1-\mathbf{x}^2)\bar{\mathbf{v}}^2/2]$, which can be further reduced to $\exp(\overline{\mathbf{v}^2}/2)$ if we set $\mathbf{x}^2 \rightarrow 1$. Setting \mathbf{x}^2 to its average value (with respect to the Gaussian PDF) is justified if $\bar{\mathbf{v}}$ and $\overline{\mathbf{v}^2}$ remain approximately constant within one imaginary-time step, which holds in the limit $\Delta\tau \rightarrow 0$. (An exception to this is when the walker $|\phi\rangle$ is in the vicinity — say, within a few time steps — of “the origin” as defined by $\langle\Psi_T|\phi\rangle = 0$. We discuss this special case, together with the characteristics of the origin and its effect on the method below.) The modified propagator in Eq. (45) can then be written approximately as

$$\int e^{-\mathbf{x}^2/2} e^{(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}} e^{\overline{\mathbf{v}^2}/2} d\mathbf{x}. \quad (50)$$

We have only included the \hat{H}_2 terms in the discussion from Eq. (40) up to now. Re-introducing \hat{H}_1 into the propagator, we obtain the complete propagator for the Hamiltonian

$$\int e^{-\mathbf{x}^2/2} \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right] \exp[(\mathbf{x}-\bar{\mathbf{v}})\cdot\hat{\mathbf{v}}] \exp\left[-\frac{\Delta\tau\hat{H}_1}{2}\right] \exp[-\Delta\tau E_L(\phi)] d\mathbf{x}, \quad (51)$$

where the term E_L has a similar form to the so-called local energy in real-space DMC

$$E_L(\phi) \equiv \frac{\langle\Psi_T|\hat{H}|\phi\rangle}{\langle\Psi_T|\phi\rangle}. \quad (52)$$

Projection with Eq. (51) will in principle lead to the ground-state wave function in the form of Eq. (38). The weight of the walker is determined by E_L , which is independent of any phase factor of the determinant.

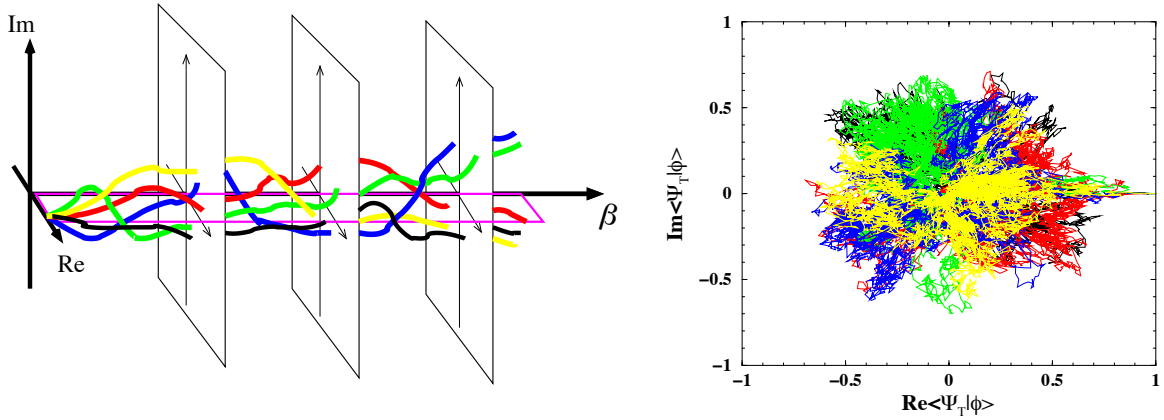


Fig. 3: Schematic illustration of the phase problem and constraints to control it. The **left panel** shows, as a function of projection time $\beta \equiv n\Delta\tau$, trajectories of 5 walkers characterized by the real (Re) and imaginary (Im) parts of their overlap with the ground-state wave function. The **right panel** shows the walker distribution integrated over imaginary time, i.e., the different frames in the left panel stacked together along β . The phase problem occurs because the initial phase “coherence” of the random walkers rapidly deteriorates with β , as they become uniformly distributed in the Re-Im -plane. The idea of the phase constraint [15] is to apply a gauge transformation such that confining the random walk in the single magenta plane (left) is a good approximation.

In the limit of an exact $|\Psi_T\rangle$, E_L is a *real* constant, and the weight of each walker remains real. The mixed estimate for the energy from Eq. (7) is phaseless

$$E_0^c = \frac{\sum_{\phi} w_{\phi} E_L(\phi)}{\sum_{\phi} w_{\phi}}. \quad (53)$$

With a general $|\Psi_T\rangle$ which is not exact, a natural approximation is to replace E_L in Eq. (51) by its real part, $\text{Re}E_L$. The same replacement is then necessary in Eq. (53).

With these replacements we arrive at a phaseless formalism for the random walks. We can summarize each step in this formalism as follows. For each random walker $|\phi\rangle$,

- (a) sample \mathbf{x} and propagate the walker to $|\phi'\rangle$

$$|\phi\rangle \rightarrow |\phi'\rangle = \exp\left[-\frac{\Delta\tau \hat{H}_1}{2}\right] \exp\left[(\mathbf{x} - \bar{\mathbf{v}}) \cdot \hat{\mathbf{v}}\right] \exp\left[-\frac{\Delta\tau \hat{H}_1}{2}\right] |\phi\rangle, \quad (54)$$

- (b) update the weight of the walker

$$w_{\phi} \rightarrow w_{\phi'} = w_{\phi} \exp\left[-\Delta\tau \cdot \text{Re}\left(E_L(\phi') + E_L(\phi)\right)/2\right]. \quad (55)$$

Walkers so generated represent the ground-state wave function with importance sampling, in the sense of Eq. (38). We have also tested a slightly different formalism based on Eq. (45), which we shall refer to as the *hybrid formalism* [15, 30]. The same choice for $\bar{\mathbf{x}}$ is used: $\bar{\mathbf{x}} = \bar{\mathbf{v}}$. The walker is propagated in the same way as in (a) above, with the shifted or force-biased propagator

in Eq. (51). In **(b)**, however, we explicitly calculate the weight factor on the left-hand side of Eq. (49) instead of using the local energy:

(b') update the weight of the walker

$$w_\phi \rightarrow w_{\phi'} = \left| \frac{\langle \Psi_T | \phi' \rangle}{\langle \Psi_T | \phi \rangle} e^{\mathbf{x} \cdot \bar{\mathbf{x}} - \bar{\mathbf{x}}^2/2} \right| \quad (56)$$

Conceptually the hybrid form is the same as that of the full local energy. However, it is in principle correct for any choice of $\bar{\mathbf{x}}$, while in the local-energy form only $\bar{\mathbf{x}} = \bar{\mathbf{v}}$ is correct. If calculating E_L is computationally costly (for example a brute-force calculation of \hat{H}_2 in a general representation in Eq. (2) costs $\mathcal{O}(N^4)$), the hybrid form can be a more efficient alternative. The Trotter errors for finite $\Delta\tau$ can be different in the two approaches [22].

This formalism is all that is needed to handle the sign problem in the case of a *real* $\hat{\mathbf{v}}$. For any $\hat{\mathbf{v}}$ the shift $\bar{\mathbf{x}}$ diverges as a walker approaches the origin in the complex plane shown in the right panel of Fig. 3, i.e., as $\langle \Psi_T | \phi' \rangle \rightarrow 0$. The effect of the divergence is to move the walker away from the origin. With a *real* $\hat{\mathbf{v}}$, $\Delta\theta = 0$ and the random walkers move only on the real axis. If they are initialized to have positive overlaps with $|\Psi_T\rangle$, $\bar{\mathbf{x}}$ will ensure that the overlaps remain positive throughout the random walk. Thus in this case our formalism above reduces to CPMC. For a general case with a complex $\hat{\mathbf{v}}$, however, the phaseless formalism alone is not sufficient to remove the phase problem. To illustrate the nature of the problem we consider the phase of $\langle \Psi_T | \phi'(\mathbf{x} - \bar{\mathbf{x}}) \rangle / \langle \Psi_T | \phi \rangle$, which we shall denote by $\Delta\theta$ and which is in general non-zero: $\Delta\theta \sim \mathcal{O}(-\mathbf{x} \cdot \text{Im}(\bar{\mathbf{x}}))$. This means that the walkers will undergo a random walk in the complex plane defined by $\langle \Psi_T | \phi' \rangle$. At large β , they will therefore populate the complex plane symmetrically, independent of their initial positions. This is illustrated in the right panel of Fig. 3, which shows $\langle \Psi_T | \phi \rangle$ for three-dimensional jellium with two electrons at $r_s = 10$ for a total projection time of $\beta = 250$. The distribution of the walkers is seen to be symmetric about the phase angle, and any signal that the walkers all real initially started with $\langle \Psi_T | \phi^{(0)} \rangle = 1$ is lost in the statistical noise. In other words, for a complex $\hat{\mathbf{v}}$, the random walk is “rotationally invariant” in the complex plane, and the divergence of $\bar{\mathbf{x}}$ is not enough to prevent the build-up of a finite density at the origin. Near the origin the local energy E_L diverges, which causes diverging fluctuations in the weights of walkers. To address this we make an additional approximation. We project the random walk to “one-dimension” and multiply the weight of each walker in each step by $\cos(\Delta\theta)$

$$w_{\phi'} = w_{\phi'} \max\{0, \cos(\Delta\theta)\}, \quad (57)$$

in addition to step **(b)** [or **(b')**] above. For this to be a good approximation, it is important that the importance sampling transformation has been done to eliminate the leading order in the overall phase of $|\phi\rangle$ in the propagator in Eq. (49).

Several alternative schemes have been tested [15, 31, 49] in place of the Eq. (57). One that seemed to work just as well was to project with the factor $\exp\{-[\text{Im}(\bar{\mathbf{x}})]^2/2\}$, which is the same as $\cos(\Delta\theta)$ in the limit of small values of $\Delta\theta$. Another was to impose $\text{Re}\langle \Psi_T | \phi' \rangle > 0$, which gave similar results, but with somewhat larger variance.

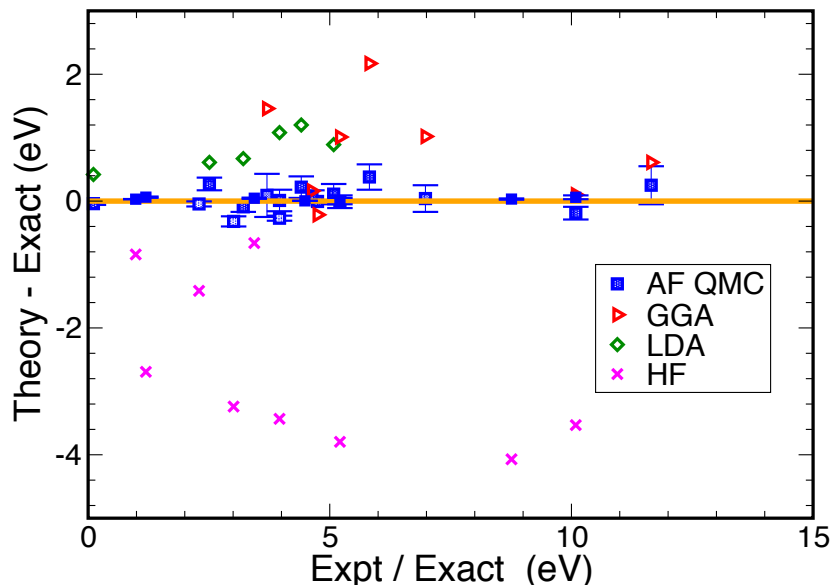


Fig. 4: Calculated binding energies [16, 27, 49, 50] of molecules compared with experimental values. The discrepancy between theory and experiment is plotted. The AFQMC is fed a trial wave function to start, which is taken directly from DFT [with either LDA or the generalized-gradient approximation (GGA) functionals] or HF. The corresponding DFT or HF results are also shown. As can be readily observed, the AFQMC results are in excellent agreement with experiment and significantly improve upon the values from DFT and HF.

4 Illustrative results

The AFQMC method, including the CPMC for real auxiliary-fields and the phaseless formalism for complex fields, has been applied to lattice models, to solids (using plane-wave basis and pseudo potentials), to molecular systems (using Gaussian basis sets), and to down-folded model Hamiltonians of real materials (using DFT orbitals as basis sets). It has also been applied in nuclear physics to treat shell models and in neutron matter calculations. Here we briefly mention a few examples from correlated-electron systems to provide an idea of what can already be done and the many opportunities for improvements and for breakthroughs.

Accuracy, or predictive power, is a key requirement of a successful paradigm for strongly correlated systems. The formulation of AFQMC with Gaussian basis sets and quantum chemistry applications have been crucial in establishing the method as a general approach that can now be applied to correlated materials, since direct comparisons can be made with high-level quantum chemistry results. Benchmark calculations to date have included close to 100 systems, from simple molecules [21, 49] and solids [22, 51, 52] to transition metal systems [53] to energy storage problems [32].

Figure 4 illustrates the results on molecules using both planewave plus pseudopotentials and Gaussian basis sets. In these calculations we have operated largely in an automated mode, inputting only the DFT or HF solutions. The method demonstrated excellent accuracy, being consistently able to correct errors in the mean-field trial wave function. In molecules, the accuracy of the phaseless AFQMC is found to be comparable to CCSD(T), the gold standard in

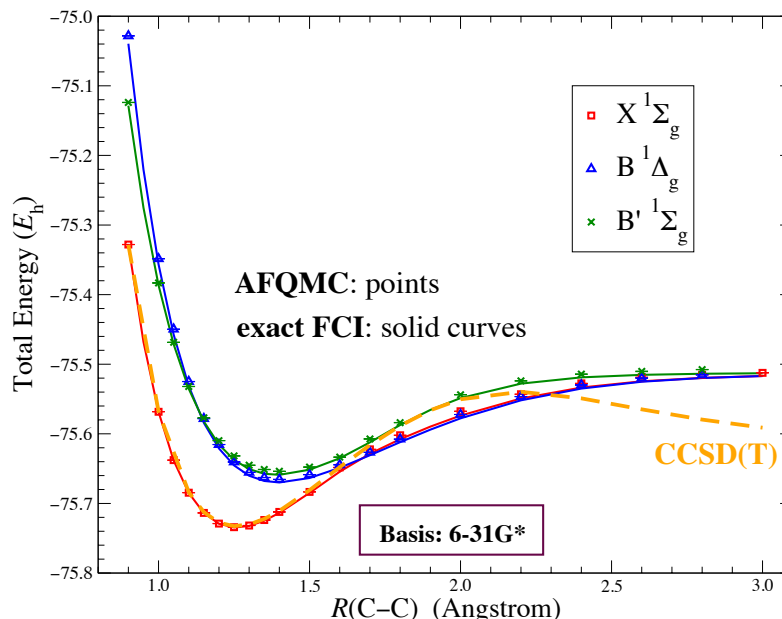


Fig. 5: Accurate calculation [58] of excited states. The seemingly simple C_2 molecule is a significant challenge to many-body methods [59,60]. Potential energy curves calculated by our QMC method (symbols with error bars) are shown for the three lowest lying singlet states in C_2 , compared with exact results [59] (solid lines) in a small basis set. RCCSD(T), the preeminent many-body quantum chemistry method, does not work well (shown in orange lines).

chemistry [54,55], near equilibrium geometry.

Bond stretching in molecules and solids mimics increased levels of multi-reference characters and particle correlation. With single Slater determinants from unrestricted HF as trial wave function, the AFQMC method generally gives better overall accuracy and a more uniform behavior than CCSD(T) in mapping the potential-energy curve [56–58], as can be seen for the example of the C_2 ground state in Fig. 5.

The AFQMC method can be used to study excited states. Prevention of collapse into the ground state and control of the fermion sign/phase problem are accomplished by the constraint using an excited state trial wave function [58]. An additional orthogonalization constraint is formulated to use virtual orbitals in solids for band structure calculations [52]. Using the C_2 molecule as a test case, we calculated the potential energy curves of the ground and two low-lying singlet excited states in Fig. 5. The trial wave function was obtained by truncating complete active space wave functions, with no further optimization. The phaseless AFQMC results using a small basis set were in good agreement with exact calculations, while those using large basis sets were in excellent agreement with experimental spectroscopic constants [58]. Applications in solids have included the calculation of the fundamental band gap in wurtzite ZnO [52].

Figure 6 illustrates an application to determine [23, 46] magnetic correlations and collective modes in the ground state of the two-dimensional repulsive Hubbard model. Using CPMC and simulating large rectangular supercells (over 1000 lattice sites), we studied the magnetic and charge properties as a function of density. At intermediate interaction strengths, an incommensurate spin density wave (SDW) state is seen, with antiferromagnetic order and essentially ho-

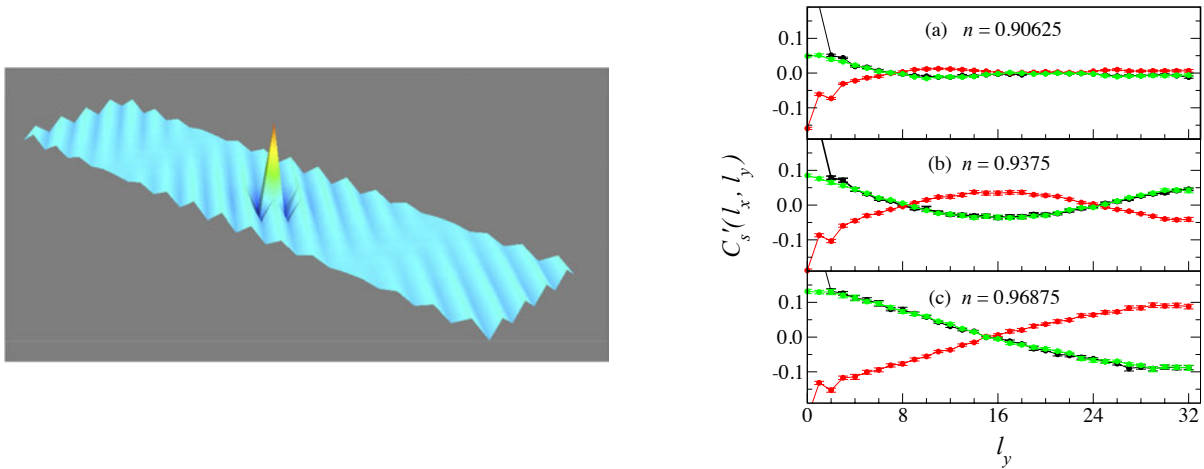


Fig. 6: Spin-density wave (SDW) states with long wavelength modulation in the 2-D Hubbard model [23]. The **left panel** is a 3-D plot of the spin-spin correlation function in the ground state of 8×32 lattice with 240 electrons ($n = 0.9375$) and $U/t = 4$. Anti-ferromagnetic order can be seen from the diagonal “lines.” Phase shifts occur at the nodal lines, across which there is a mismatch of the diagonal lines. The **right panel** examines the doping dependence. The staggered spin-spin correlation function is plotted at three different densities. Calculations are done using a free-electron trial wave function. The long wavelength modulation and nodal lines (two colors crossing) are clear. As doping is increased, the wavelength of the modulating wave decreases, as does the amplitude of the SDW.

homogeneous charge correlations. The wavelength of the collective mode decreases with doping $(1 - (M_{\uparrow} + M_{\downarrow})/N)$, as does its magnitude. The SDW order vanishes beyond a critical doping. As the interaction is increased, the holes go from a wavelike (delocalized) to a particlelike (localized) state, and charge ordering develops which eventually evolves into stripe states [61]. This work [23] advanced our understanding of the magnetic properties and provided predictions for optical lattice experiments.

5 Concluding remarks

We have discussed ground-state auxiliary-field-based methods for correlated-electron systems. A formalism has been outlined that allows for a systematic understanding of the origin of the sign/phase problem, and a platform on which to build methods whose computational cost scales with system size at a low power, and whose accuracy reaches or surpasses the best general many-body approaches. The AFQMC methods can be used to study both model systems and realistic Hamiltonians. Applications have included atoms, molecules, solids, and correlated-electron models, recently including down-folded Hamiltonians.

The original form of the AFQMC method allows essentially exact calculations of ground-state and finite-temperature equilibrium properties of interacting many-fermion systems. Their effectiveness, however, was severely limited by the sign problem, which prevents calculations at large system sizes or low temperatures. Often in the study of correlated models in condensed

matter or cold atoms in optical lattices, the comment “But there is a sign problem” is made to imply “So QMC calculations cannot be used here.” It is hoped that the methods we have discussed will change this paradigm. Calculations indeed can be done in systems where a sign problem is present, and often accurate results are within reach by these calculations.

As a QMC method, this framework automatically accounts for particle permutations and allows easy computations of both diagonal and off-diagonal expectations, as well as imaginary-time correlations. It has much versatility as a general method, offering, for example, the possibility to treat heavier elements and spin-orbit coupling, and the potential to embed the AFQMC naturally and seamlessly in a calculation at the mean-field level. While much work is needed to study various forms of HS transformations and trial functions $|\psi_T\rangle$, and to understand the subtleties of the method and the constraints because of the non-orthogonal and over-complete nature of the Slater determinant space involved, we expect the method and the concept brought forth here to see many applications, and to significantly enhance the applicability of quantum simulations in interacting fermion systems. There are many opportunities both for further development and improvement of this framework, and for applications in a variety of problems.

6 Acknowledgments

I am indebted to numerous colleagues and students, whose interactions and collaborations I have immensely enjoyed and benefitted from. Length restriction does not allow the full list here, but I would like to thank W. Al-Saidi, J. Carlson, C.-C. Chang, H. Krakauer, F. Ma, W. Purwanto, and H. Shi. The work covered in these notes would not have been possible without their contributions. Support from the National Science Foundation (NSF), the Department of Energy (DOE), and the Army Research Office (ARO) is gratefully acknowledged. Computing was done on the Oak Ridge Leadership Computing Facilities via an INCITE award, on Blue Waters at UIUC supported by NSF PRAC, and on CPD at William & Mary.

Appendices

A A few basics of Monte Carlo techniques

We list a few key elements from standard Monte Carlo (MC) techniques which are important to our discussions on QMC. For an introduction on MC methods, see, e.g., Ref. [42].

MC methods are often used to compute many-dimensional integrals of the form

$$G = \frac{\int_{\Omega_0} f(\mathbf{x})g(\mathbf{x})d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x})d\mathbf{x}}, \quad (58)$$

where \mathbf{x} is a vector in a many-dimensional space and Ω_0 is a domain in this space. We will assume that $f(\mathbf{x}) \geq 0$ on Ω_0 and that it is normalizable, i.e., the denominator is finite. A familiar example of the integral in Eq. (58) comes from classical statistical physics, where $f(\mathbf{x})$ is the Boltzmann distribution.

To compute G by MC, we *sample* \mathbf{x} from a probability density function (PDF) proportional to $f(\mathbf{x})$: $\bar{f}(\mathbf{x}) \equiv f(\mathbf{x}) / \int_{\Omega_0} f(\mathbf{x})d\mathbf{x}$. This means to generate a sequence $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots\}$ so that the probability that any \mathbf{x}_i is in the sub-domain $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ is

$$\text{Prob}\{\mathbf{x}_i \in (\mathbf{x}, \mathbf{x} + d\mathbf{x})\} = \bar{f}(\mathbf{x})d\mathbf{x}. \quad (59)$$

There are different techniques to sample a many-dimensional function $f(\mathbf{x})$. The most general and perhaps most commonly used technique to sample $f(\mathbf{x})$, i.e., the PDF $\bar{f}(\mathbf{x})$, is the Metropolis algorithm, which creates a Markov-chain random walk [42] in \mathbf{x} -space whose equilibrium distribution is the desired function. We will also use a branching random walk, in which case there can be a weight w_i associated with each sampled \mathbf{x}_i . (In Metropolis, $w_i = 1$.) The MC samples provide a formal representation of f

$$f(\mathbf{x}) \propto \sum_i w_i \delta(\mathbf{x} - \mathbf{x}_i). \quad (60)$$

Given \mathcal{M} independent samples from $f(\mathbf{x})$, the integral in Eq. (58) is estimated by

$$G_{\mathcal{M}} = \frac{\sum_{i=1}^{\mathcal{M}} w_i g(\mathbf{x}_i)}{\sum_{i=1}^{\mathcal{M}} w_i} \quad (61)$$

The error in the estimate decays algebraically with the number of samples: $|G_{\mathcal{M}} - G| \propto 1/\sqrt{\mathcal{M}}$. Using the results above, we can compute

$$G' = \frac{\int_{\Omega_0} f(\mathbf{x})g(\mathbf{x})h(\mathbf{x})d\mathbf{x}}{\int_{\Omega_0} f(\mathbf{x})h(\mathbf{x})d\mathbf{x}}, \quad (62)$$

if the function $h(\mathbf{x})$ is such that both the numerator and denominator exist. Formally

$$G'_{\mathcal{M}} = \frac{\sum_{i=1}^{\mathcal{M}} w_i g(\mathbf{x}_i)h(\mathbf{x}_i)}{\sum_{i=1}^{\mathcal{M}} w_i h(\mathbf{x}_i)}, \quad (63)$$

although, as we will see, difficulties arise when $h(\mathbf{x})$ can change sign and is rapidly oscillating. Integral equations are another main area of applications of MC methods. For example [42], the integral equation

$$\Psi'(\mathbf{x}) = \int_{\Omega_0} K(\mathbf{x}, \mathbf{y}) w(\mathbf{y}) \Psi(\mathbf{y}) d\mathbf{y}, \quad (64)$$

can be viewed in terms of a *random walk* if it has the following properties: $\Psi(\mathbf{y})$ and $\Psi'(\mathbf{x})$ can be viewed as PDF's (in the sense of f in Eq. (58)), $w(\mathbf{y}) \geq 0$, and $K(\mathbf{x}, \mathbf{y})$ is a PDF for \mathbf{x} conditional on \mathbf{y} . Then, given an ensemble $\{\mathbf{y}_i\}$ sampling $\Psi(\mathbf{y})$, the following two steps will allow us to generate an ensemble that samples $\Psi'(\mathbf{x})$. First an absorption/branching process is applied to each \mathbf{y}_i according to $w(\mathbf{y}_i)$. For example, we can make $\text{int}(w(\mathbf{y}_i) + \xi)$ copies of \mathbf{y}_i , where ξ is a uniform random number on $(0, 1)$. Second we randomly walk each new \mathbf{y}_j to an \mathbf{x}_j by sampling the PDF $K(\mathbf{x}_j, \mathbf{y}_j)$. The resulting $\{\mathbf{x}_j\}$ are MC samples of $\Psi'(\mathbf{x})$. An example of this is the one-dimensional integral equation

$$\Psi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x-y)^2} \sqrt{2} e^{-y^2/2} \Psi(y) dy, \quad (65)$$

which has a solution $\Psi(x) = e^{-x^2/2}$. The random walks, starting from an arbitrary distribution, will iteratively converge to a distribution sampling $\Psi(x)$.

References

- [1] See, e.g., N.-P. Ong and R. Bhatt (Ed.):
More is different: Fifty years of condensed matter physics (Princeton Univ. Press, 2001)
- [2] R.M. Martin: *Electronic Structure: Basic theory and practical methods*
(Cambridge University Press, Cambridge, UK, 2004)
- [3] See, e.g., A. Szabo and N.S. Ostlund: *Modern quantum chemistry*
(McGraw-Hill, New York, 1989)
- [4] W. Kohn, Rev. Mod. Phys. **71**, 1253 (1999) and references therein
- [5] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)
- [6] S.Y. Savrasov, G. Kotliar, and E. Abrahams, Nature **410**, 793 (2001)
- [7] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, Phys. Rev. Lett. **87**, 067205 (2001)
- [8] K. Held, A.K. McMahan, and R.T. Scalettar, Phys. Rev. Lett. **87**, 276404 (2001)
- [9] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)
- [10] M.H. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974)
- [11] W.M.C. Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001)
and references therein
- [12] D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995) and references therein
- [13] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D **24**, 2278 (1981)
- [14] G. Sugiyama and S.E. Koonin, Ann. Phys. (NY) **168**, 1 (1986)
- [15] S. Zhang and H. Krakauer, Phys. Rev. Lett. **90**, 136401 (2003)
- [16] K.P. Esler, J. Kim, D.M. Ceperley, W. Purwanto, E.J. Walter, H. Krakauer, S. Zhang,
P.R.C. Kent, R.G. Hennig, C. Umrigar, M. Bajdich, J. Kolorenc, L. Mitas, and A. Srinivasan,
Journal of Physics: Conference Series **125**, 012057 (2008)
- [17] K.E. Schmidt and M.H. Kalos, in K. Binder (ed.):
Applications of the Monte Carlo Method in Statistical Physics (Springer, 1984)
- [18] E.Y. Loh Jr., J.E. Gubernatis, R.T. Scalettar, S.R. White, D.J. Scalapino, and R. Sugar,
Phys. Rev. B **41**, 9301 (1990)
- [19] S. Zhang: *Constrained path Monte Carlo for fermions* (cond-mat/9909090)
in M.P. Nightingale and C.J. Umrigar (eds.):
Quantum Monte Carlo Methods in Physics and Chemistry (Kluwer Dordrecht, 1999)

- [20] S. Zhang, J. Carlson, and J.E. Gubernatis, Phys. Rev. B **55**, 7464 (1997)
- [21] W.A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **124**, 224101 (2006)
- [22] W. Purwanto, H. Krakauer, and S. Zhang, Phys. Rev. B **80**, 214116 (2009)
- [23] C.-C. Chang and S. Zhang, Phys. Rev. Lett. **104**, 116402 (2010)
- [24] S. Zhang: in D. Sénéchal, A.-M. Tremblay, and C. Bourbonnais (eds.): *Theoretical Methods for Strongly Correlated Electrons* (Springer, 2003), and references therein;
<http://physics.wm.edu/~shiwei>
- [25] S.B. Fahy and D.R. Hamann, Phys. Rev. Lett. **65**, 3437 (1990)
- [26] J. Carlson, J.E. Gubernatis, G. Ortiz, and S. Zhang, Phys. Rev. B **59**, 12788 (1999)
- [27] M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E.J. Walter, Phys. Rev. B **75**, 245123 (2007)
- [28] S. Zhang, Phys. Rev. Lett. **83**, 2777 (1999)
- [29] B.M. Rubenstein, S. Zhang, and D.R. Reichman, Phys. Rev. A **86**, 053606 (2012)
- [30] W. Purwanto and S. Zhang, Phys. Rev. E **70**, 056702 (2004)
- [31] W. Purwanto and S. Zhang, Phys. Rev. A **72**, 053610 (2005)
- [32] W. Purwanto, H. Krakauer, Y. Virgus, and S. Zhang, J. Chem. Phys. **135**, 164105 (2011)
- [33] S. Zhang and D.M. Ceperley, Phys. Rev. Lett. **100**, 236404 (2008)
- [34] D.R. Hamann and S.B. Fahy, Phys. Rev. B **41**, 11352 (1990)
- [35] J. Carlson, S. Gandolfi, K E. Schmidt, and S. Zhang, Phys. Rev. A **84**, 061602 (2011)
- [36] R.L. Stratonovich, Sov. Phys. Dokl. **2**, 416 (1958)
- [37] J.E. Hirsch, Phys. Rev. B **28**, 4059 (1983)
- [38] W.A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **126**, 194105 (2007)
- [39] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Perseus Books, Reading, Massachusetts, 1998)
- [40] R. Baer, M. Head-Gordon, and D. Neuhauser, J. Chem. Phys. **109**, 6219 (1998)
- [41] H. Shi and S. Zhang, Phys. Rev. B, arXiv:1307.2147 (2013)
- [42] M.H. Kalos and P.A. Whitlock: *Monte Carlo methods* Vol. I (Wiley, 1986)

- [43] S. Sorella, S. Baroni, R. Car, and M. Parrinello, *Europhys. Lett.* **8**, 663 (1989)
- [44] S. Zhang and M.H. Kalos, *Phys. Rev. Lett.* **67**, 3074 (1991)
- [45] J.E. Hirsch, *Phys. Rev. B* **31**, 4403 (1985)
- [46] C.-C. Chang and S. Zhang, *Phys. Rev. B* **78**, 165101 (2008)
- [47] S.R. White, D.J. Scalapino, R.L. Sugar, E.Y. Loh, J.E. Gubernatis, and R.T. Scalettar, *Phys. Rev. B* **40**, 506 (1989)
- [48] J.W. Moskowitz, K.E. Schmidt, M.A. Lee, and M.H. Kalos, *J. Chem. Phys.* **77**, 349 (1982)
- [49] S. Zhang, H. Krakauer, W.A. Al-Saidi, and M. Suewattana, *Comput. Phys. Commun.* **169**, 394 (2005)
- [50] W.A. Al-Saidi, H. Krakauer, and S. Zhang, *Phys. Rev. B* **73**, 075103 (2006)
- [51] H. Kwee, S. Zhang, and H. Krakauer, *Phys. Rev. Lett.* **100**, 126404 (2008)
- [52] F. Ma, S. Zhang, and H. Krakauer, *New J. of Phys.* (arXiv:1211.4635) (2013)
- [53] Y. Virgus, W. Purwanto, H. Krakauer, and S. Zhang, *Phys. Rev. B* **86**, 241406 (2012)
- [54] R.J. Bartlett and M. Musiał, *Rev. Mod. Phys.* **79**, 291 (2007)
- [55] T.D. Crawford and H.F. Schaefer III, *Reviews in Computational Chemistry* **14**, 33 (2000)
- [56] W.A. Al-Saidi, S. Zhang, and H. Krakauer, *J. Chem. Phys.* **127**, 144101 (2007)
- [57] W. Purwanto, W.A. Al-Saidi, H. Krakauer, and S. Zhang, *J. Chem. Phys.* **128**, 114309 (2008)
- [58] W. Purwanto, S. Zhang, and H. Krakauer, *J. Chem. Phys.* **130**, 094107 (2009)
- [59] M.L. Abrams and C.D. Sherrill, *J. Chem. Phys.* **121**, 9211 (2004)
- [60] C.J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R.G. Hennig, *Phys. Rev. Lett.* **98**, 110201 (2007)
- [61] S.R. White and D.J. Scalapino, *Phys. Rev. Lett.* **91**, 136403 (2003)

16 DMRG: Ground States, Time Evolution, and Spectral Functions

Ulrich Schollwöck

Ludwig-Maximilians-Universität München

Theresienstraße 37, 80333 Munich, Germany

Contents

1	DMRG: A young adult	2
2	Matrix product states	2
3	Matrix product operators	11
4	Normalization and compression	13
5	Time-evolution: tDMRG, TEBD, tMPS	15
6	Overlaps and expectation values	18
7	Finite-temperature simulations	21
8	Ground states with MPS: DMRG	23
9	Constructing the MPO representation of a Hamiltonian	26
10	Dynamical DMRG	27
11	Outlook: DMRG in two dimensions	29

1 DMRG: A young adult

On November 9, 1992, Physical Review Letters published a paper entitled “Density Matrix Formulation for Quantum Renormalization Groups” by Steven R. White [1]. After introducing a new algorithm for calculating the low-lying states and their observables of one-dimensional lattice models, the paper presented impressively precise results for both the spin- $\frac{1}{2}$ and spin-1 Heisenberg antiferromagnet. It concludes: “This new formulation appears extremely powerful and versatile, and we believe it will become the leading numerical method for 1D systems; and eventually will become useful for higher dimensions as well.” I must admit that I am very surprised that these very confident sentences made it past the referees; but this was just as well: the successes of the *density matrix renormalization group* (DMRG) within the last two decades have, if anything, far exceeded the hopes Steve White might have harboured for his new algorithm at that time.

As all youths do, DMRG underwent puberty, developing a completely new personality: around 2004, when it was 12 years old, the (much older) realization [2, 3] that DMRG is closely linked to a special quantum state class, so-called *matrix product states* (MPS), suddenly spawned a number of algorithmic extensions, which drastically enhanced the reach of DMRG (to name but the first few: [4–10]). These algorithmic extensions could all be expressed in both DMRG and MPS language, so many practitioners at first preferred to stay with the old way of speaking. However, at least in my view it is now abundantly clear that formulating DMRG in the language of MPS is notationally much cleaner and conceptually much more adequate, so I will present the entire lecture in this way. The price to pay is to get used to a notation which will definitely be unfamiliar to a physicist with standard training, as opposed to the DMRG language. But it is worth the effort! For those readers who develop a deeper interest, an overview of DMRG in the old language, with some focus on fields of application, is given by [11], whereas a very technical, but hopefully thorough introduction into the structure and manipulation of MPS can be found in [12]. A more conceptual orientation which presents the story as seen by quantum information theory is found in [13].

2 Matrix product states

Let us consider a quantum system that lives on L lattice sites with d local states $\{\sigma_i\}$ on each site $i \in \{1, 2, \dots, L\}$. A good example to think of would be interacting spins- $\frac{1}{2}$ where the local states are $|\uparrow\rangle, |\downarrow\rangle$ and $d = 2$. The lattice may at this point be of arbitrary dimension, we just have to give an ordering to the site labels. In view of what is to come, one may of course think of a one-dimensional chain with sites 1 through L .

Pure states are then defined on the L^d -dimensional Hilbert space

$$\mathcal{H} = \bigotimes_{i=1}^L \mathcal{H}_i \quad \mathcal{H}_i = \{|1_i\rangle, \dots, |d_i\rangle\} \quad (1)$$

and the most general state reads

$$|\psi\rangle = \sum_{\sigma_1, \dots, \sigma_L} c^{\sigma_1 \dots \sigma_L} |\sigma_1 \dots \sigma_L\rangle. \quad (2)$$

In the following, we will often abbreviate as $\{\sigma\} = \sigma_1 \dots \sigma_L$. The usual problem of numerical simulations is that the number of state coefficients $c^{\{\sigma\}}$ grows exponentially with system size L . A standard first approximation in order to reduce the exponential number of coefficients, the so-called *mean-field approximation*, consists in factorizing the state coefficients as

$$c^{\sigma_1 \dots \sigma_L} = c^{\sigma_1} \cdot c^{\sigma_2} \cdot \dots \cdot c^{\sigma_L}. \quad (3)$$

Instead of d^L coefficients, we now have dL coefficients (in the special case that we can assume translational invariance of the state, the number reduces even more drastically to d). The motivation of this ansatz dates back to the molecular field theory of Weiss (1907), where the orientation of small elementary magnets (spin was not known at that time) is calculated by assuming they are exposed to an external magnetic field and an additional effective magnetic field, which self-consistently models the interaction with all other elementary magnets. Successful as mean-field theories have been over the decades (the BCS theory of superconductivity (1957), for example, is structurally a mean-field theory of Cooper pairs), they do not capture the essential feature of quantum physics: *entanglement*.

What makes quantum mechanics fundamentally different from classical physics is the different nature of the states: whereas they are points in phase space in classical physics, they are rays in Hilbert space in quantum mechanics; moreover, the Hilbert space of a many-particle system is given by the tensor product of the single-particle Hilbert spaces. To consider the most simple example, take 2 spins- $\frac{1}{2}$. Then $\mathcal{H}_i = \{|\uparrow_i\rangle, |\downarrow_i\rangle\}$ and $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. The combination of the superposition principle (implied by states living in Hilbert space) and the tensor product structure means that the most general state is

$$|\psi\rangle = c^{\uparrow\uparrow} |\uparrow\uparrow\rangle + c^{\uparrow\downarrow} |\uparrow\downarrow\rangle + c^{\downarrow\uparrow} |\downarrow\uparrow\rangle + c^{\downarrow\downarrow} |\downarrow\downarrow\rangle, \quad (4)$$

subject merely to the normalization condition. It is now very easy to show that not every state of this form factorizes, i.e. $c^{\uparrow\downarrow} = c^{\uparrow} c^{\downarrow}$ and so forth; to see this, just consider the singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}} |\uparrow\downarrow\rangle - \frac{1}{\sqrt{2}} |\downarrow\uparrow\rangle. \quad (5)$$

States that factorize are called *product states*, whereas all others are called *entangled states*. The importance of entangled states is of course given by the fact that they carry non-local information and superclassical correlations, both of which are essential to quantum physics [14–16]. How can we generalize the product state by Eq. (3) to describe (at least certain) entangled states, while remaining numerically convenient? As general states are sums over products of local states, we are led to think about matrices replacing the scalars of Eq. (3); we are then looking at the sum of product states; they can be entangled. So the most simple generalization would be

$$c^{\sigma_1} \cdot c^{\sigma_2} \cdot \dots \cdot c^{\sigma_L} \rightarrow M^{\sigma_1} \cdot M^{\sigma_2} \cdot \dots \cdot M^{\sigma_L}, \quad (6)$$

where all M^{σ_i} are (2×2) -matrices, except on sites 1 and L , where they must be (1×2) and (2×1) row and column vectors respectively, such that the matrix product yields a scalar. In fact, the famous Affleck-Kennedy-Lieb-Tasaki (AKLT) model has a ground state that can be cast in exactly this form and contains a wealth of non-trivial physics [17, 18]. In fact, this simple ansatz can also be used as a variational ansatz for the ground state of entire classes of Hamiltonians giving quite deep insights into their physics; see, for example, [19–21]. Of course, (2×2) -matrices are only of limited descriptive power given the wealth of quantum states.

We therefore consider the following generalization, which is a generic matrix product state:

$$|\psi\rangle = \sum_{\sigma_1, \dots, \sigma_L} M^{\sigma_1} M^{\sigma_2} \dots M^{\sigma_L} |\sigma_1 \sigma_2 \dots \sigma_L\rangle, \quad (7)$$

where at each site we introduce d matrices M^{σ_i} , which therefore depend on the local state $|\sigma_i\rangle$. The dimensions of the matrices are $(1 \times D_1), (D_1 \times D_2), \dots, (D_{L-2} \times D_{L-1}), (D_{L-1} \times 1)$, with equal row and column indices of matrices associated to neighboring sites such that the matrix product can be carried out, with the very first and last dimension 1, to yield a scalar.

A given state $|\psi\rangle$ does *not* have a unique decomposition into matrices M^{σ_i} : to see this consider an arbitrary, but invertible matrix X of dimension $(D_i \times D_i)$. Then the matrix product state does not change under the insertion of $XX^{-1} = 1$ between matrices M^{σ_i} and $M^{\sigma_{i+1}}$, which implies a gauge transformation

$$M^{\sigma_i} \rightarrow M^{\sigma_i} X \quad M^{\sigma_{i+1}} \rightarrow X^{-1} M^{\sigma_{i+1}}. \quad (8)$$

Later on, we will exploit this gauge degree of freedom to bring MPS into a particularly efficient form for practical use.

Why is this state class so interesting? There are, in my view, five reasons for this.

1. Any quantum state can be represented as an MPS, although the representation may be numerically inefficient. Nevertheless, it is therefore a mathematical structure of general interest.
2. There is a hierarchy of MPS in the sense that states with low entanglement can be represented more efficiently (using smaller matrices) than highly entangled states. So-called area laws reveal that these low-entanglement states are particularly important for low-temperature quantum physics. This makes MPS useful in practice. For the link between MPS and the entanglement of quantum states, I refer to the lecture of Jens Eisert.
3. They emerge naturally in the context of renormalization group schemes, connecting the DMRG framework to more conventional RG schemes.
4. They can be manipulated easily and efficiently: this concerns the application of operators ranging from local creation or annihilation operators through Hamiltonians to time evolution operators, the evaluation of overlaps and expectation values.
5. They can be searched efficiently: which state, given \hat{H} , has the lowest energy among a given class of MPS? This is the nucleus of a variational method (none other but DMRG).

To prove the first statement, we need either of two matrix decompositions from linear algebra, the QR or the singular value decomposition (SVD); for an excellent first introduction into these techniques of numerical linear algebra, I recommend [22]. In this presentation, I will use the SVD because it will cover all numerical needs we are going to have; however, in numerical practice, the QR factorisation is much faster and should be used whenever possible instead of the SVD (I will indicate these occasions). For an arbitrary $(m \times n)$ -matrix A , we have, with $k = \min(m, n)$, the following decomposition

$$A = USV^\dagger, \quad (9)$$

where the matrices have the following special properties:

- U is $(m \times k)$ -dimensional and consists of orthonormal columns, i.e., $U^\dagger U = I$; if $m = k$, then $UU^\dagger = I$ too, and U is unitary.
- S is $(k \times k)$ -dimensional and diagonal. The entries on the diagonal are called the *singular values* s_i and are real and non-negative, $s_i \geq 0$. The number $r \leq k$ of strictly positive singular values is equal to the rank of A . All texts and computer codes assume that singular values are sorted in descending order, $s_1 \geq s_2 \geq s_3 \geq \dots$
- V^\dagger is $(k \times n)$ -dimensional and consists of orthonormal rows, i.e., $V^\dagger V = I$; if $k = n$, then $VV^\dagger = I$ too, and V is unitary.

“Old-fashioned” DMRG makes heavy use of the *eigenvalue decomposition (EVD)* $AU = U\Lambda$ for quadratic (and hermitean) A , with Λ a diagonal matrix with real (but not necessarily non-negative) eigenvalues λ_i on the diagonal and U a unitary matrix whose column vectors $|u_i\rangle$ can be taken to be an orthonormal basis. Here, incidentally, I have introduced a very useful notation in linear algebra, namely that matrices are also read as sets of column vectors, $U = [|u_1\rangle |u_2\rangle \dots]$, or of row vectors (which is used less frequently). SVD and EVD are however closely connected: if $A = USV^\dagger$, then

$$A^\dagger A = VSU^\dagger USV^\dagger = VS^2V^\dagger \Rightarrow (A^\dagger A)V = VS^2 \quad (10)$$

and similarly

$$AA^\dagger = USV^\dagger V SU^\dagger = US^2U^\dagger \Rightarrow (AA^\dagger)U = US^2. \quad (11)$$

Comparing to the EVD, this means that the singular values squared are the eigenvalues of both $A^\dagger A$ and AA^\dagger , and the respective eigenvectors are the columns of U and V respectively. This allows to translate between SVD based procedures in MPS-based algorithms and EVD based procedures in classic DMRG.

Before concluding this tiny excursion into the truly rich properties of the SVD, let me mention that it is also behind a very important decomposition of quantum states, the so-called *Schmidt decomposition*. Its key importance is that it allows for a direct readout of the entanglement properties of a state: consider a bipartition of the “universe” AB (in our case: the L sites) into

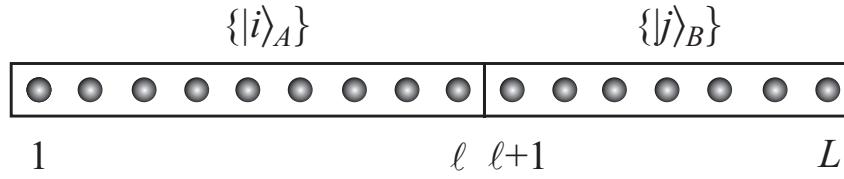


Fig. 1: Bipartitioning a quantum system in blocks A and B with block-local orthonormal bases.

subsystems A and B (in our case e.g. sites 1 through ℓ and $\ell + 1$ through L); see Fig. 1. Then the most general pure quantum state $|\psi\rangle$ reads

$$|\psi\rangle = \sum_{i=1}^{\dim \mathcal{H}_A} \sum_{j=1}^{\dim \mathcal{H}_B} \psi_{ij} |i\rangle_A |j\rangle_B, \quad (12)$$

where the $\{|i\rangle_A\}$ and $\{|j\rangle_B\}$ form orthonormal bases of subsystems A and B respectively; in practice, these bases will usually be product bases resulting from local orthonormal bases like $\{|\sigma_1 \dots \sigma_\ell\rangle\}$. Interpreting the ψ_{ij} as the entries of a rectangular matrix Ψ , an SVD of Ψ gives $\Psi = USV^\dagger$ and we can rewrite our state in the *Schmidt decomposed* form

$$|\psi\rangle = \sum_{\alpha=1}^r s_\alpha |\alpha\rangle_A |\alpha\rangle_B \quad (13)$$

with s_α the (rank) r non-vanishing singular values of Ψ and

$$|\alpha\rangle_A = \sum_{i=1}^{\dim \mathcal{H}_A} U_{i\alpha} |i\rangle_A \quad |\alpha\rangle_B = \sum_{j=1}^{\dim \mathcal{H}_B} V_{j\alpha}^* |j\rangle_B \quad (14)$$

It is crucial to note that, due to the properties of U and V , the $\{|\alpha\rangle_A\}$ and $\{|\alpha\rangle_B\}$ form orthonormal sets respectively, because a pitfall of the MPS world is that one encounters state decompositions that look like Eq. (13), but where orthonormality does *not* hold; hence they are not Schmidt decompositions. The physics of the subsystems A and B is encoded by the reduced density operators

$$\hat{\rho}_A = \text{tr}_B |\psi\rangle\langle\psi| = \sum_{\alpha=1}^r s_\alpha^2 |\alpha\rangle_A \langle\alpha| \quad \hat{\rho}_B = \text{tr}_A |\psi\rangle\langle\psi| = \sum_{\alpha=1}^r s_\alpha^2 |\alpha\rangle_B \langle\alpha| \quad (15)$$

which the Schmidt decomposition allows to be read off in eigenrepresentation. This allows interesting observations like the identity of the non-vanishing eigenvalues of the reduced density operators on A and B, even if A and B are very different: the difference only shows up in the eigenvectors. This, by the way, is a neat illustration of the SVD/EVD link.

If we quantify entanglement by the von Neumann entropy of entanglement, then the entanglement between subsystems A and B given a quantum state $|\psi\rangle$ is obtained by the (conventional) von Neumann entropy of either of the subsystems A or B:

$$S_{A|B}(|\psi\rangle) = -\text{tr}_A \hat{\rho}_A \ln \hat{\rho}_A = -\text{tr}_B \hat{\rho}_B \ln \hat{\rho}_B = -\sum_{\alpha=1}^r s_\alpha^2 \ln s_\alpha^2. \quad (16)$$

We can use this formula and work out the entanglement of the two states already encountered, the general product state and the singlet state of two spins.

In the case of the product state, we consider some bipartitioning into sites 1 through ℓ (subsystem A) and $\ell + 1$ through L (subsystem B). We can then write

$$|\psi\rangle = |\alpha\rangle_A |\alpha\rangle_B \quad \text{with} \quad |\alpha\rangle_{A,B} = \sum_{\{\sigma_{A,B}\}} c^{\sigma_{A,B}} |\sigma_{A,B}\rangle, \quad (17)$$

where $\sigma_A \equiv \sigma_1, \dots, \sigma_\ell$ and similarly σ_B . The state is right away in a Schmidt representation, and the eigenvalue spectrum of the reduced density operators is $(1, 0, 0, \dots)$; inserting this in the von Neumann entropy formula and using $0 \ln 0 = \lim_{\epsilon \rightarrow 0^+} \epsilon \ln \epsilon = 0$, we find that this state is *unentangled*. As the product state is an MPS with matrix dimensions 1, we can state that an MPS is unentangled if and only if it can be represented exactly by an MPS with matrix dimensions 1. (I should add “for a given basis”, because entanglement has the unnerving and deeply unphysical property that it is not invariant under global basis transformations: for a simple example, consider the triplet states at $S^z = \pm 1$ which are product states, but which can be rotated by a global basis transformation into the $S^z = 0$ triplet state which is entangled.) The take-home message for us is that, as already mentioned, $D = 1$ MPS are classical (in the sense of unentangled) states and that life gets quantum mechanical (interesting) only from $D = 2$ onwards.

In the case of the singlet state, the reduced density operator obtained by a partial trace is already in diagonal form, $\hat{\rho}_A = \hat{\rho}_B = \text{diag}(\frac{1}{2}, \frac{1}{2})$. Entanglement then is given by $-2 \cdot \frac{1}{2} \ln \frac{1}{2} = \ln 2$. (In statistical physics texts the base of the logarithm is e ; in quantum information texts the base of the logarithm is 2 as befits the world of qubits, such that entanglement here would be exactly 1.) The interest of this result is that, using Lagrangian multipliers, one can easily show that this is the maximum entanglement any two-spin state could have and that it is achieved if and only if the reduced density operators are *maximally mixed*, i.e., all eigenvalues are identical. This makes sense as entanglement is a measure of the amount of non-local information in a quantum system: a maximally mixed reduced density operator implies that the amount of local information is minimal. We can generalize the result to a reduced density operator of dimension D : the maximally mixed density operator then has eigenvalues D^{-1} , and the maximal entanglement is given by $-D \cdot D^{-1} \ln D^{-1} = \ln D$.

After this lengthy detour, I am now going to demonstrate that every state can in principle be represented as an MPS. This is done by successively peeling off site after site (say, starting with 1, but the reverse procedure is also possible). Consider $c^{\sigma_1 \sigma_2 \dots \sigma_L}$, the coefficients of the d^L -dimensional state vector and *reshape* them into a $(d \times d^{L-1})$ -dimensional matrix, which is then SV decomposed:

$$c^{\sigma_1 \sigma_2 \dots \sigma_L} \rightarrow \Psi_{\sigma_1, \sigma_2 \dots \sigma_L} = \sum_{a_1} U_{\sigma_1, a_1} S_{a_1, a_1} V_{a_1, \sigma_2 \dots \sigma_L}^\dagger. \quad (18)$$

The matrix U is now sliced into d row vectors A^{σ_1} , which we interpret as $(1 \times d)$ matrices (d being the maximum rank possible):

$$U_{\sigma_1, a_1} \rightarrow \{A^{\sigma_1}\} \quad \text{with} \quad A_{1, a_1}^{\sigma_1} = U_{\sigma_1, a_1}. \quad (19)$$

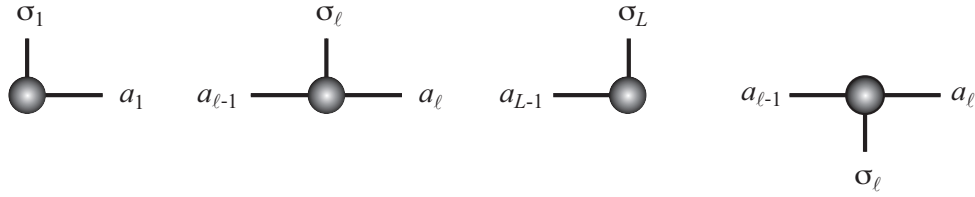


Fig. 2: Building blocks of an MPS: matrices for first and last sites, as well as bulk sites. Physical indices point vertically, matrix indices horizontally. Complex-conjugation mirrors matrices along the horizontal axis.

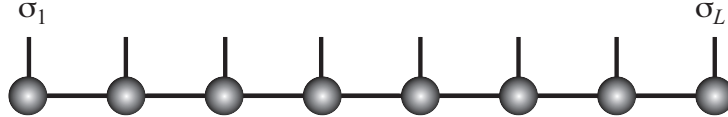


Fig. 3: Graphical representation of an MPS: all connected lines between building blocks are contracted over.

If we lump together $c^{a_1\sigma_2\sigma_3\ldots\sigma_L} = S_{a_1,a_1} V_{a_1,\sigma_2\ldots\sigma_L}^\dagger$, we have as a first step

$$c^{\sigma_1\sigma_2\ldots\sigma_L} = \sum_{a_1} A_{1,a_1}^{\sigma_1} c^{a_1\sigma_2\sigma_3\ldots\sigma_L}. \quad (20)$$

In a second step (and all others will be the same), we will again use a sequence of reshaping and a SVD,

$$c^{a_1\sigma_2\sigma_3\ldots\sigma_L} \rightarrow \Psi_{a_1\sigma_2,\sigma_3\ldots\sigma_L} = \sum_{a_2} U_{a_1\sigma_2,a_2} S_{a_2,a_2} V_{a_2,\sigma_3\ldots\sigma_L}^\dagger. \quad (21)$$

U is now sliced “horizontally” into d matrices A^{σ_2} of dimension $(d \times d^2)$, where

$$A_{a_1,a_2}^{\sigma_2} = U_{a_1\sigma_2,a_2}, \quad (22)$$

and if S and V are again lumped together to form a new $c^{a_2\sigma_3\sigma_4\ldots\sigma_L}$, the state coefficients read after the second step

$$c^{\sigma_1\sigma_2\ldots\sigma_L} = \sum_{a_1,a_2} A_{1,a_1}^{\sigma_1} A_{a_1,a_2}^{\sigma_2} c^{a_2\sigma_3\sigma_4\ldots\sigma_L}. \quad (23)$$

It is easy to see that upon continuation the coefficients can be represented as a product of matrices as in the definition of an MPS. After half the chain, matrix dimensions which first grow as $(1 \times d)$, $(d \times d^2)$, and so on, will shrink again because in SVD $\min(m, n)$ sets the dimension. The largest matrix dimension is $d^{L/2}$ (L assumed even), hence again exponential in L and it seems nothing has been gained: we still need to approximate, i.e. replace matrices by smaller ones (say, of some maximum dimension $D \sim 1000$) while minimizing the loss of accuracy.

As should be obvious by now, the manipulation of MPS is extremely index-heavy. Fortunately, there is by now a generally accepted graphical representation of MPS (Fig. 2) and their building blocks (up to one degree of freedom, which is nothing compared to the confusing notational wealth surrounding Green’s functions or the sign conventions in general relativity). Each matrix is drawn as a dot with two horizontal and one vertical line sticking out; the two horizontal lines

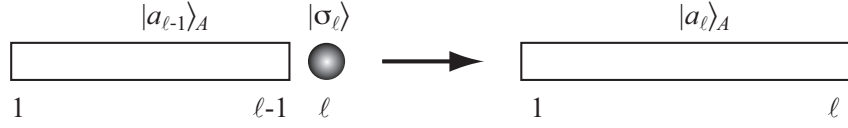


Fig. 4: Block growth of classic DMRG.

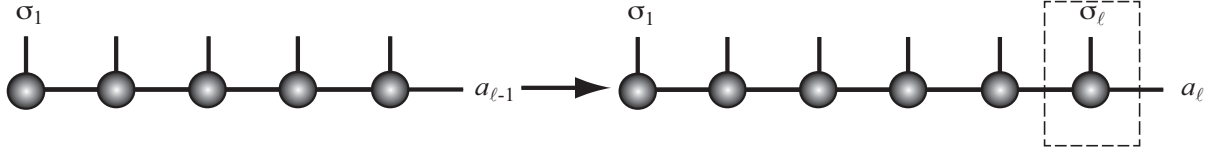


Fig. 5: Block growth of classic DMRG represented in the MPS framework.

correspond to the row and column indices and the vertical line to the physical index labeling the matrix. The degree of freedom is whether the vertical line points up or down. I let it point up. If the index only takes the value 1, hence is a dummy index, the corresponding line is dropped (this concerns first and last sites). If one wants to represent the complex-conjugated (not adjoint!) matrix $M^{\sigma*}$, an object we will need for the bra $\langle\psi|$, the direction of the vertical line is inverted, pointing down in my convention. For building an entire state or other more complicated structures, there is a single rule: all connected lines are contracted, i.e., the connected objects are multiplied and the joint index is summed over; for two matrices this is obviously a matrix multiplication. A generic MPS as in Eq. (7) would then look like a comb (Fig. 3).

Some readers of these notes may be familiar with DMRG in the original notation of White, using blocks (of sites) and sites. Let me put this notation into a more general context, that of a generic 1D renormalization scheme: consider a semi-infinite lattice of sites with d local degrees of freedom. If we group the first ℓ sites into a block, a state on the block has d^ℓ coefficients, which is exponentially large. If we decide that we build such blocks *iteratively*, starting from a “block” containing just one site and adding site after site (Fig. 4), then we have to devise a *decimation* scheme of discarding basis states to avoid the exponential growth $d \rightarrow d^2 \rightarrow d^3 \rightarrow \dots$ and to keep the number of states manageable. If we decide to keep D states for a block description and assume that we have a basis $\{|a_{\ell-1}\rangle\}$ for the block of length $\ell - 1$ and add site ℓ (local states $\{|\sigma_\ell\rangle\}$), then the D states building the (incomplete!) basis of the block of length ℓ will read

$$|a_\ell\rangle = \sum_{a_{\ell-1}, \sigma_\ell} \langle a_{\ell-1}, \sigma_\ell | a_\ell \rangle |a_{\ell-1}\rangle |\sigma_\ell\rangle \equiv \sum_{a_{\ell-1}, \sigma_\ell} M_{a_{\ell-1}, a_\ell}^{\sigma_\ell} |a_{\ell-1}\rangle |\sigma_\ell\rangle, \quad (24)$$

where we have reorganized the expansion coefficients into d matrices M labelled by the local state; the entries are given by $M_{a_{\ell-1}, a_\ell}^{\sigma_\ell} = \langle a_{\ell-1}, \sigma_\ell | a_\ell \rangle$ and connect states on the smaller block and the larger block. The matrix entries therefore encode the decimation scheme, which can now be represented as in Fig. 5. The reorganization of the coefficients is of interest because we can iterate the scheme such that

$$|a_\ell\rangle = \sum_{\sigma_1, \dots, \sigma_\ell} (M^{\sigma_1} M^{\sigma_2} \dots M^{\sigma_\ell})_{1, a_\ell} |\sigma_1 \sigma_2 \dots \sigma_\ell\rangle. \quad (25)$$



Fig. 6: Graphical representation of left and right normalization conditions (left and right part of the figure). The single lines represent identities.

The sums over the matrix row and column indices was simply absorbed into a compact matrix multiplication notation, and structurally the basis states look like MPS. As we have specified no decimation procedure, we can draw several conclusions: (i) DMRG is a method that grows blocks using decimation yields states in the MPS format (with subtle modifications that can be ignored here); (ii) any RG scheme in 1D that can be characterized by an iterative growth and decimation scheme yields states in the MPS format; this holds in particular for Wilson's Numerical Renormalization Group (NRG) [23,24] which is the method of choice for the famous Kondo problem. In fact, as a historical remark, the failure of Wilson's NRG for general strongly correlated problems in 1D was what motivated White's work in 1992; the underlying MPS structure finally allowed to understand the connection between NRG and DMRG [25].

I already mentioned in the beginning that there is a gauge degree of freedom in MPS. Both in the general decomposition and in the block growth procedure we have unwittingly chosen a gauge which will turn out to be extremely useful in practice. In the block growth procedure both $\{|a_{\ell-1}\rangle\}$ and $\{|a_\ell\rangle\}$ form orthonormal sets, respectively. Therefore

$$\begin{aligned} \delta_{a'_\ell, a_\ell} &= \langle a'_\ell | a_\ell \rangle = \sum_{a'_{\ell-1} \sigma'_\ell a_{\ell-1} \sigma_\ell} M_{a'_{\ell-1}, a'_\ell}^{\sigma'_\ell *} M_{a_{\ell-1}, a_\ell}^{\sigma_\ell} \langle a'_{\ell-1} \sigma'_\ell | a_{\ell-1} \sigma_\ell \rangle \\ &= \sum_{a_{\ell-1} \sigma_\ell} M_{a_{\ell-1}, a'_\ell}^{\sigma_\ell *} M_{a_{\ell-1}, a_\ell}^{\sigma_\ell} = \sum_{\sigma_\ell} (M^{\sigma_\ell \dagger} M^{\sigma_\ell})_{a'_\ell, a_\ell} \end{aligned}$$

or

$$I = \sum_{\sigma_\ell} M^{\sigma_\ell \dagger} M^{\sigma_\ell} \equiv \sum_{\sigma_\ell} A^{\sigma_\ell \dagger} A^{\sigma_\ell}. \quad (26)$$

Matrices that obey this relationship are called *left-normalized* and will be denominated by A ; exactly the same property follows from the general state decomposition from the column orthonormality $U^\dagger U = I$. If one builds blocks from the right, adding sites at the left end, or if one carries out an SVD on a general state starting from the right, i.e. at site L , one obtains similarly *right-normalized* matrices B with

$$I = \sum_{\sigma_\ell} B^{\sigma_\ell} B^{\sigma_\ell \dagger}. \quad (27)$$

I will refer to MPS that consist entirely of A -matrices as *left-canonical*, MPS that consist entirely of B -matrices I will call *right-canonical*. A third class, which is the most important one in numerical practice, is called *mixed-canonical* and has structure $AAAAAMB BBBB BBBBBB$, with one matrix without special normalization property sandwiched in between. We will discuss further below how to convert between all three representations. Graphically, the normalization conditions are represented as in Fig. 6.

3 Matrix product operators

Given that quantum mechanics works by applying operators (unitary operators for time evolution, projection operators in measurements, ...) to states, it is quite surprising that the systematic representation of operators in a generalization of the MPS scheme did not really start until quite recently (see e.g. [26]). The most general operator on our L sites reads

$$\hat{O} = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} c^{\sigma_1 \dots \sigma_L, \sigma'_1 \dots \sigma'_L} |\sigma_1 \dots \sigma_L\rangle \langle \sigma'_1 \dots \sigma'_L|, \quad (28)$$

where the primed variables label the ingoing state the operator acts on and the unprimed variables the outgoing state. If we reshuffle the indices to group states on the same site, we have

$$c^{\sigma_1 \dots \sigma_L, \sigma'_1 \dots \sigma'_L} \rightarrow c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \dots \sigma_L \sigma'_L} \quad (29)$$

and a “mean-field approximation” to the operator would read

$$c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \dots \sigma_L \sigma'_L} \rightarrow c^{\sigma_1 \sigma'_1} \cdot c^{\sigma_2 \sigma'_2} \cdot \dots \cdot c^{\sigma_L \sigma'_L}. \quad (30)$$

While there is no physical reason why this should be in any ways a good approximation, in fact this is an *exact* representation of many operators: consider the operator \hat{S}_i^z acting on a spin on site i . As this operator is in fact defined only on \mathcal{H}_i , a pedantic notation would be

$$\hat{S}_i^z \rightarrow \hat{I}_1 \otimes \hat{I}_2 \otimes \dots \otimes \hat{S}_i^z \otimes \dots \otimes \hat{I}_L \quad (31)$$

and the coefficients would read

$$c^{\sigma_1 \sigma'_1 \sigma_2 \sigma'_2 \dots \sigma_L \sigma'_L} = \delta_{\sigma_1, \sigma'_1} \cdot \delta_{\sigma_2, \sigma'_2} \cdot \dots \cdot (\hat{S}_i^z)_{\sigma_i, \sigma'_i} \cdot \dots \cdot \delta_{\sigma_L, \sigma'_L}. \quad (32)$$

So, if we introduce *matrix product operators* (MPO) as a straightforward generalization of the MPS notation as

$$\hat{O} = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} M^{\sigma_1 \sigma'_1} M^{\sigma_2 \sigma'_2} \dots M^{\sigma_L \sigma'_L} |\sigma_1 \dots \sigma_L\rangle \langle \sigma'_1 \dots \sigma'_L| \quad (33)$$

with the usual rules for the matrix dimensions, the above operator would be encoded simply by scalar (i.e., $D = 1$) matrices, and this would also hold for operators as used for n -point correlators such as $\hat{S}_i^+ \hat{S}_j^-$. Any more complicated operator can also be turned into an MPO: reconsider the construction of an MPS for an arbitrary state, and group indices $(\sigma_1 \sigma'_1)(\sigma_2 \sigma'_2) \dots$. Of course, the question arises whether this neat result will generalize in the sense that more complicated operators like a Hamiltonian \hat{H} still find an *exact* representation with *small* D . As opposed to MPS of interest which will usually involve an approximation, we will find this to be true for many Hamiltonians of interest.

What happens if we apply an MPO \hat{O} to an MPS $|\psi\rangle$? Assuming that the MPS is formed by matrices M^{σ_i} and the MPO by matrices $N^{\sigma_i \sigma'_i}$ it is a simple calculation to show that the new state $\hat{O}|\psi\rangle$ is again an MPS, with matrices \tilde{M}^{σ_i} that have entries

$$\tilde{M}_{(ab), (a'b')}^{\sigma_i} = \sum_{\sigma'_i} N_{aa'}^{\sigma_i \sigma'_i} M_{bb'}^{\sigma'_i}. \quad (34)$$

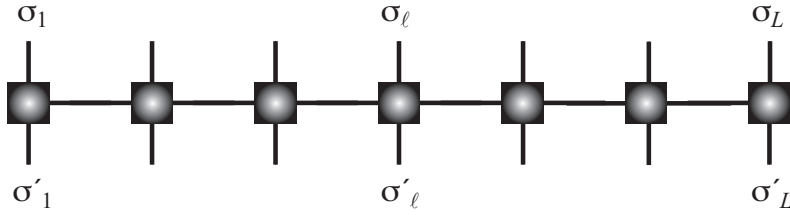


Fig. 7: Graphical representation of an MPO: the vertical lines sticking out to the bottom represent ingoing physical states, the lines sticking out to the top outgoing physical states.

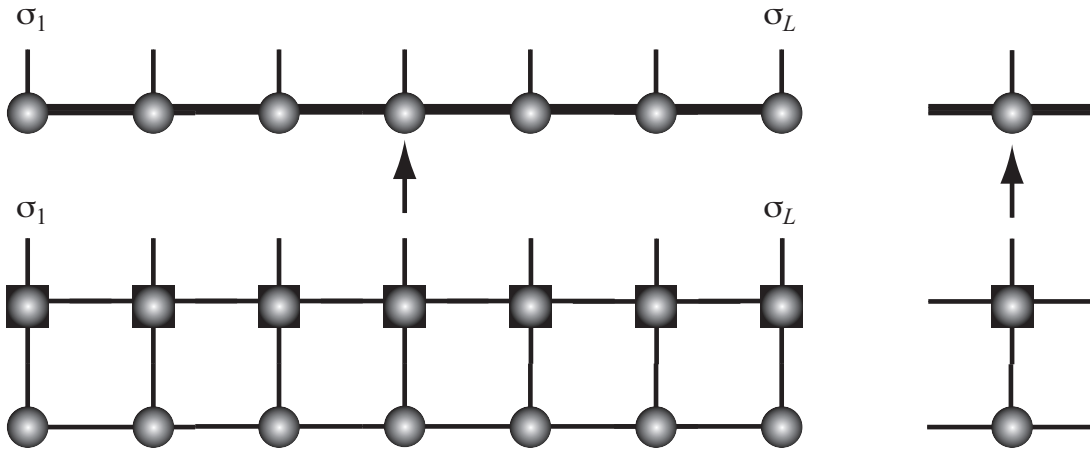


Fig. 8: Graphical representation of the application of an MPO to an MPS: all connected lines between building blocks are contracted over. A new MPS results, with matrix dimensions being the product of the original matrix dimensions.

Numerically, this operation can be implemented very efficiently; the important observation is that the dimensions of the new matrices are given by the *product* of the old matrix dimensions. This is potentially disastrous, as at least MPS dimensions can become very large for some desired quality of approximation. But as most Hamiltonians in MPO form have dimensions less than 10, the new MPS may still be barely manageable; nevertheless this observation indicates that we need a *compression procedure* such that an MPS with undesirably large (unmanageable) matrix dimensions can be approximated optimally by an MPS with smaller matrix dimensions at some loss of accuracy.

Graphically, an MPO (Fig. 7) is represented in analogy to an MPS, with *two* vertical legs sticking out, corresponding to the “ingoing” (line pointing down) and “outgoing” state (line pointing up). The MPO times MPS gives MPS rule finds a simple graphical representation as in Fig. 8. Lines sticking out in the same direction can be unified into a single one, with the rule that the dimension of the unified line is the product of the original dimensions.

4 Normalization and compression

As we have seen, the dimensions of MPS will usually grow as we proceed with calculations, making them potentially useless if we do not counter this growth. How can we compress an MPS with minimal loss of information? Let us assume (don't worry that the assumption looks a bit artificial) that we have a quantum state in the following MPS representation,

$$|\psi\rangle = \sum_{\{\sigma\}} A^{\sigma_1} A^{\sigma_2} \dots A^{\sigma_\ell} M^{\sigma_{\ell+1}} B^{\sigma_{\ell+2}} \dots B^{\sigma_L} |\sigma_1 \dots \sigma_L\rangle \quad (35)$$

and the row and column dimensions of $M^{\sigma_{\ell+1}}$ (and the adjacent column dimension of A^{σ_ℓ} and row dimension of $B^{\sigma_{\ell+2}}$) are too big and we want to reduce them with minimum of loss of accuracy in the state description. To achieve this, we stack the d matrices $M^{\sigma_{\ell+1}}$ columnwise into a single matrix, i.e.

$$M_{a_\ell, \sigma_{\ell+1} a_{\ell+1}} = M_{a_\ell, a_{\ell+1}}^{\sigma_{\ell+1}} \quad (36)$$

and carry out an SVD on the new matrix as $M = U S V^\dagger$. If we absorb U it into $A^{\sigma_\ell} \leftarrow A^{\sigma_\ell} U$, this retains the left-normalization due to $U^\dagger U = I$ and corresponds to a basis transformation of the (reduced) orthonormal block basis for block A formed from sites 1 through ℓ ,

$$|a_\ell\rangle_A := \sum_{\sigma_1, \dots, \sigma_\ell} (A^{\sigma_1} \dots A^{\sigma_\ell})_{1, a_\ell} |\sigma_1 \dots \sigma_\ell\rangle. \quad (37)$$

Similarly, we have a new orthonormal block basis on the block formed by sites $\ell + 1$ through L by slicing as

$$B_{a_\ell, a_{\ell+1}}^{\sigma_{\ell+1}} = V_{a_\ell, \sigma_{\ell+1} a_{\ell+1}}^\dagger, \quad (38)$$

where $B^{\sigma_{\ell+1}}$ are right-normalized and

$$|a_\ell\rangle_B := \sum_{\sigma_{\ell+1}, \dots, \sigma_L} (B^{\sigma_{\ell+1}} \dots B^{\sigma_L})_{a_\ell, 1} |\sigma_{\ell+1} \dots \sigma_L\rangle \quad (39)$$

form an orthonormal set. Identifying $s_{a_\ell} = S_{a_\ell, a_\ell}$, we therefore have a Schmidt decomposition of $|\psi\rangle$ as

$$|\psi\rangle = \sum_{a_\ell} s_{a_\ell} |a_\ell\rangle_A |a_\ell\rangle_B. \quad (40)$$

Hence (and this is why the mixed-canonical representation is so important, ensuring orthonormal bases on A and B) a mixed-state representation can be turned into a Schmidt decomposition and vice versa. The Schmidt decomposition, on the other hand, allows to read off the correct compression strategy: the s_a^2 corresponding to the statistical weights in the reduced density operators, the optimal truncation is given by *retaining those pairs of Schmidt states that have the maximum Schmidt coefficients*. If we can afford matrix dimensions D , and assume ordering by descending singular values, we simply cut down the column dimension of A^{σ_ℓ} to D and similarly the row dimension of $B^{\sigma_{\ell+1}}$, and both dimensions of S .

The problem with this procedure is that it only works at the seam between the left-normalized and right-normalized parts of the state, but in general we have to truncate everywhere. Our

procedure immediately indicates how to remedy this: if (after truncation) we multiply S to the left, $M^{\sigma_\ell} \leftarrow A^{\sigma_\ell} S$, the matrices on site ℓ will lose their normalization property and the state will have the form

$$|\psi\rangle = \sum_{\{\sigma\}} A^{\sigma_1} A^{\sigma_2} \dots A^{\sigma_{\ell-1}} M^{\sigma_\ell} B^{\sigma_{\ell+1}} \dots B^{\sigma_L} |\sigma_1 \dots \sigma_L\rangle. \quad (41)$$

Compared to our initial state, the seam between left- and right-normalized states has shifted by one to the left, and we can continue our truncation procedure now one site to the left. This, of course, would also have been the case if we had not truncated at all.

This allows us to define two strategies: in order to bring *any* MPS into form $AAAA\dots$ or $BBBB\dots$, we start either from site 1 (for A) or site L (for B) and work our way through the chain by a sequence of SVDs *without* truncations (in such a case, faster QR decompositions do as well). In the previous paragraphs, we have just seen how a step to the left generates a B . The fact that we had A -matrices to the left there only mattered for the truncation. Bringing any MPS into left-canonical or right-canonical form I refer to as *normalization* (indeed, as one can see easily, in the very last step of the procedure, a scalar survives, which is nothing but the norm of the state, so one can use it as a “conventional” normalization procedure $\langle\psi|\psi\rangle \stackrel{!}{=} 1$ as well). Partially right-normalizing a left-normalized MPS generates a mixed-normalized MPS.

As a second strategy for compressing an MPS to some acceptable matrix dimension, we take a state that is e.g. in form $AAAAA\dots$ (which can be achieved by the first strategy) and move through all mixed-canonical representations from the right, truncating along the way. Here, the QR factorization, which does not give access to the singular values, is not useful.

At this point it is now easy to see for which types of states MPS can yield good approximations: the quality of truncations depends on how quickly the singular values s_a of the Schmidt decompositions (or the statistical weights of the reduced density operators) decay with a : if they decay rapidly, the truncated statistical weight is negligible, and in practice truncated weights for ground states are often of the order 10^{-10} or even less if we keep matrix dimensions $D \sim 1000$. Usually, we do not know the spectrum of the reduced density operators; as entanglement is derived from this spectrum, it contains similar (but much less) information: a state with a rapidly decaying spectrum has lower entanglement than a state with slowly decaying spectrum. Hence, the feasibility of the representation of a state by an MPS rests on its entanglement properties, MPS being a low-entanglement representation. For more details I refer to Jens Eisert’s lecture in this volume.

Looking more closely, one realizes that this compression scheme cannot be absolutely optimal: while it is optimal at each step, the compression at some site depends on the outcome of earlier compressions on other sites, but not on the compression that will happen later. Hence there is an informational asymmetry, which usually is not so important because the amount of compression at each site is small, as just mentioned. In cases where this is a problem, there exists an alternative variational technique which is strictly optimal, see [12].

5 Time-evolution: tDMRG, TEBD, tMPS

Traditionally, the exposition of DMRG starts with explaining the ground state algorithm, DMRG proper, which for a given Hamiltonian \hat{H} looks for its ground state within, as it turns out, the space of MPS. This reflects the historical course of events: time-dependent DMRG [4–6, 8] (with the variants of TEBD and tMPS, but this is all very much the same) was invented 12 years after ground state DMRG. We will reverse this sequence, as time-dependent DMRG is much easier to explain and also implement *and* can be used, albeit in a quite inefficient way, to find the ground state of a given Hamiltonian.

We restrict our attention to time-independent Hamiltonians \hat{H} ; this captures a large number of the problems encountered in practice. As all more important time-evolution schemes currently in use consider small (“infinitesimal”) time steps, time-dependent Hamiltonians can be modeled by a sequence of Hamiltonians that change after each small time step.

Assume we have an initial state $|\psi(0)\rangle$ in MPS form; such a state can be constructed by hand (in simple cases like a Néel state, which is just a $D = 1$ MPS) or is obtained by some other MPS calculation, e.g. as the ground state of some (other) Hamiltonian (otherwise there would be no non-trivial dynamics) – this is the typical setup in ultra cold atom experiments where non-equilibrium dynamics is generated by Hamiltonian quenches, i.e. abrupt changes in Hamiltonian parameters. In the case of coherent evolution, the state at time t is given by

$$|\psi(t)\rangle = e^{-i\hat{H}t}|\psi(0)\rangle. \quad (42)$$

If we manage to give $e^{-i\hat{H}t}$ an MPO representation, the problem would be solved within the MPS framework, as applying an MPO to an MPS yields a new MPS. There are several issues with this idea: (i) no one knows how to do this exactly in an efficient form on a classical computer, let alone a piece of paper; (ii) usually we are interested in the entire evolution, i.e. the state for an entire sequence of times; (iii) if the dimension of the resulting MPO is too large, the resulting MPS will not be numerically manageable. While problems (i) and (ii) can be resolved at some cost in accuracy, we will see (iii) to be a fundamental issue.

One approach, that is also historically the first, is to *Trotterize* time evolution. This is a well-known analytical scheme first used in quantum field theory, and then first applied in numerical physics by Suzuki [27] to quantum Monte Carlo schemes. Assume we want to calculate $|\psi(t)\rangle$ for times in $[0, T]$. We split interval length T into $N \rightarrow \infty$ time steps $\tau \rightarrow 0$ with $N\tau = T$. In numerical practice, τ will of course be solidly finite, e.g. $\tau = 0.01$ when a typical time scale of the problem is 1. Let us also assume that our Hamiltonian consists only of at most nearest-neighbor terms, like $\mathbf{S}_i \cdot \mathbf{S}_{i+1}$ in the Heisenberg model or the hopping term $\sum_{\sigma} (\hat{c}_{i\sigma}^{\dagger} \hat{c}_{i+1\sigma} + \text{h.c.})$ in the Hubbard model. Then we can split the Hamiltonian in nearest-neighbour terms, $\hat{H} = \sum_{i=1}^{L-1} \hat{h}_i$, where in the case of the Heisenberg model $\hat{h}_i = \mathbf{S}_i \cdot \mathbf{S}_{i+1}$. On-site terms like the Hubbard interaction are “distributed” across the two \hat{h} that share a site and counted only half, e.g. $U \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$ enters as $(U/2) \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$ in both \hat{h}_{i-1} and \hat{h}_i . Watch out for first and last sites!

Then we can rewrite the evolution operator as

$$e^{-i\hat{H}T} = \prod_{i=1}^N e^{-i\hat{H}\tau} = \prod_{k=1}^N e^{-i\sum_{i=1}^{L-1} \hat{h}_i \tau} \stackrel{!}{=} \prod_{k=1}^N \prod_{i=1}^{L-1} e^{-i\hat{h}_i \tau}. \quad (43)$$

Not only do we now have access to all times $\tau, 2\tau, \dots$, we now also only have to calculate the (infinitesimal) time-evolution on *two sites*, $e^{-i\hat{h}_i \tau}$. This is a $(d^2 \times d^2)$ matrix obtained by diagonalizing the $(d^2 \times d^2)$ matrix representation H_i of \hat{h}_i and exponentiating it,

$$H_i U = U \Lambda \Rightarrow H_i = U \Lambda U^\dagger \Rightarrow e^{-iH_i \tau} = U e^{-i\Lambda \tau} U^\dagger = U \cdot \text{diag}(e^{-i\lambda_1 \tau}, e^{-i\lambda_2 \tau}, \dots) \cdot U^\dagger, \quad (44)$$

which is easily implemented using standard diagonalization and matrix multiplication routines. So the fundamental building block can be calculated, but there is a catch, indicated by the exclamation mark on the last identity in (43). This factorization does *not* work, due to Glauber's formula,

$$e^{\hat{A}+\hat{B}} = e^{\hat{A}} e^{\hat{B}} e^{\frac{1}{2}[\hat{A}, \hat{B}]}, \quad (45)$$

which only holds under some restrictions on \hat{A} and \hat{B} (there may be further terms), but the decisive point is that if two operators do not commute, the exponential of their sum will *not* factorize in general. As our operators \hat{A} and \hat{B} scale as τ , however, in our case the commutator will scale with τ^2 . Hence, in the limit $\tau \rightarrow 0$, factorization becomes exact; for finite τ , the error will scale as τ^2 and can therefore be excellently extrapolated to the exact $\tau \rightarrow 0$ limit. This decomposition of the evolution operator is called the *first-order Trotter decomposition*, because after $N = T/\tau$ time steps, the accumulated error will be of order τ . Higher-order decompositions are available (fourth-order decompositions being the most popular, with an error of order τ^4 ; for details, I refer to the literature), but for our purpose of explaining the method first-order serves perfectly well.

As bond evolution operators commute if they do not share a site, it is customary to split the Hamiltonians into odd and even bonds as

$$\hat{H} = \hat{H}_{\text{odd}} + \hat{H}_{\text{even}}; \quad \hat{H}_{\text{odd}} = \sum_i \hat{h}_{2i-1}, \quad \hat{H}_{\text{even}} = \sum_i \hat{h}_{2i} \quad (46)$$

and arrange time evolution as

$$e^{-i\hat{H}T} = e^{-i\hat{H}_{\text{even}}\tau} e^{-i\hat{H}_{\text{odd}}\tau}; \quad e^{-i\hat{H}_{\text{even}}\tau} = \prod_i e^{-i\hat{h}_{2i}\tau}, \quad e^{-i\hat{H}_{\text{odd}}\tau} = \prod_i e^{-i\hat{h}_{2i-1}\tau}. \quad (47)$$

Assuming we know the two-site time-evolution MPOs, we can represent the Trotter-decomposed MPO for the global time evolution as in Fig. 9.

The two-site time-evolution MPO can easily be derived by our SVD decomposition procedure developed for MPS. It is easy to calculate $U^{\sigma_1 \sigma_2, \sigma'_1 \sigma'_2} = \langle \sigma_1 \sigma_2 | e^{-i\hat{h}_1 \tau} | \sigma'_1 \sigma'_2 \rangle$. Then we proceed by reshuffling of indices and SVD as

$$U^{\sigma_1 \sigma_2, \sigma'_1 \sigma'_2} = \bar{U}_{\sigma_1 \sigma'_1, \sigma_2 \sigma'_2} \stackrel{SVD}{=} \sum_b W_{\sigma_1 \sigma'_1, b} S_{b, b} W_{b, \sigma_2 \sigma'_2} = \sum_b M_{1, b}^{\sigma_1 \sigma'_1} M_{b, 1}^{\sigma_2 \sigma'_2}$$

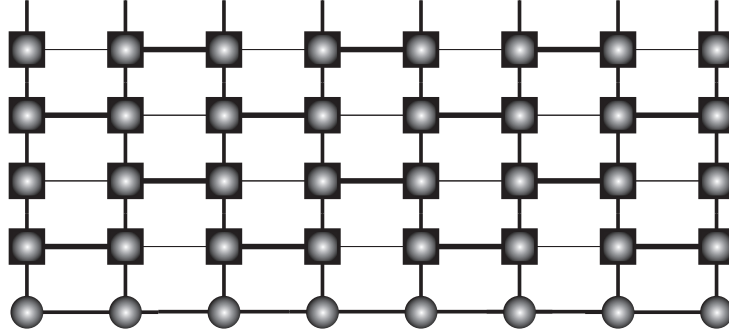


Fig. 9: MPO representation of the Trotterized time-evolution operator $e^{-i\hat{H}T}$; every second line corresponds to an odd-bond or even-bond infinitesimal time-evolution; at the bottom, there is the state that is evolved in time contracted into the MPO.

which is just an MPO where I have absorbed two factors of $\sqrt{S_{b,b}}$ into the left and right matrices. As the original matrix is of dimension $(d^2 \times d^2)$, the MPO dimension is $D_W = d^2$; more precisely, the matrices for odd bond evolutions have dimensions $(1 \times d^2)$, $(d^2 \times 1)$, $(1 \times d^2)$, \dots and vice versa for even bonds. After one Trotter step, MPS matrix dimensions will therefore have grown everywhere by a factor d^2 . This means that after each time step, we will have to compress the MPS, and the time evolution algorithm takes a very simple form:

1. Apply infinitesimal Trotter time step in MPO form to MPS to obtain $|\psi(t)\rangle \rightarrow |\psi(t + \tau)\rangle$
2. Compress $|\psi(t + \tau)\rangle$
3. Continue with the next time step

In principle, it looks as if this procedure could be continued forever. The Trotter decomposition error, which is $O(\tau^2)$ per time step will accumulate as $t/\tau O(\tau^2)$, hence grow linearly in time. It can be made arbitrarily small by sending $\tau \rightarrow \infty$ or, more efficiently, using higher-order Trotter decompositions where the error will scale as $O(\tau^3)$ or even $O(\tau^5)$ in the most frequently used approaches. The problem rests in the compression; compression as such is not a problem if we can describe the state $|\psi(t + \tau)\rangle$ with the same accuracy as $|\psi(t)\rangle$ by an MPS of given dimension D . This is however not the case: building on the Lieb-Robinson theorem [28], it has been observed that entanglement in a time-evolving quantum state will grow up to linearly in time, $S(t) \leq S(0) + \nu t$, with some constant ν , and that this linear bound is met in “global quenches” where the non-equilibrium is generated by a sudden change of the Hamiltonian operator. As we have seen, a reduced density operator of dimension D (hence an MPS of dimension D) can encode at most entanglement $S = \ln D$; hence we have to increase matrix sizes in the MPS at worst *exponentially* as time evolves to maintain the same accuracy. Time-dependent simulations therefore hit an exponential wall after some time. No one has found a solution for that problem yet. While observing thermalization is therefore excluded, in many situations of interest, ν is small enough that we can see the physics of interest; as a few examples, consider spin-charge separation [29], the propagation of bosonic correlations in ultracold atom gases [30], the

relaxation of a density wave in strongly interacting ultracold atoms [31, 32], but there are many more.

Let me conclude this section by briefly discussing the calculation of ground states using *imaginary* time evolution. Quite generally, starting from a random state $|\psi\rangle = \sum_n c_n |n\rangle$, with eigenstates $\hat{H}|n\rangle = E_n|n\rangle$, $E_0 \leq E_1 \leq E_2 \leq \dots$,

$$\begin{aligned} \lim_{\beta \rightarrow \infty} e^{-\beta \hat{H}} |\psi\rangle &= \lim_{\beta \rightarrow \infty} \sum_n e^{-\beta E_n} c_n |n\rangle = \lim_{\beta \rightarrow \infty} e^{-\beta E_0} (c_0 |0\rangle + \sum_{n>0} e^{-\beta(E_n - E_0)} c_n |n\rangle) \\ &= \lim_{\beta \rightarrow \infty} e^{-\beta E_0} c_0 |0\rangle, \end{aligned}$$

where for simplicity I have also assumed that the ground state is non-degenerate. We see that if the random starting state has some overlap with the ground state, this contribution will survive longest in the $\beta \rightarrow \infty$ limit. Of course, except for the unlikely case that $E_0 = 0$, this surviving contribution will either diverge or decay exponentially. But as we have to truncate after each infinitesimal (imaginary) time step anyways, we can use this occasion to normalize the state at every step, yielding the ground state $|0\rangle$ in the large- β limit.

Numerically, imaginary time evolution is more benign than real time evolution, as any errors made numerically are exponentially suppressed by further applications of $e^{-\tau \hat{H}}$, whereas it is a hallmark of a unitary (real time) evolution that errors are with us to stay and will only be compounded with further errors. However, the procedure is slow compared to a direct variational ground state search in MPS space, which is classic DMRG, and will be discussed below.

6 Overlaps and expectation values

Calculating time-evolving states as such is of course of little interest; in the end, we want to calculate how much a state is changing from the original state (i.e. the overlap $\langle \psi(t) | \psi(0) \rangle$) or, even more frequently, how some observable evolves in time, e.g. $\langle S_i^z(t) \rangle = \langle \psi(t) | \hat{S}_i^z | \psi(t) \rangle$. Let us first focus on the overlap, because the latter calculation will be a simple generalization. Mathematically, representing $|\psi\rangle$ by matrices M and $|\phi\rangle$ by matrices \tilde{M} , we have

$$\langle \phi | \psi \rangle = \sum_{\{\sigma\}} \sum_{\{\sigma'\}} \langle \{\sigma'\} | \tilde{M}^{\sigma'_1*} \dots \tilde{M}^{\sigma'_L*} M^{\sigma_1} \dots M^{\sigma_L} | \{\sigma\} \rangle = \sum_{\{\sigma\}} \tilde{M}^{\sigma_1*} \dots \tilde{M}^{\sigma_L*} M^{\sigma_1} \dots M^{\sigma_L}. \quad (48)$$

Graphically, this can be represented as in Fig. 10. In which order are we to carry out the contractions? At first sight, this may seem a detail, but it is not: imagine we contract all the horizontal lines first. Each contraction is a matrix-matrix multiplication that costs $O(D^3)$ assuming matrix size D . Hence there are $O(2LD^3)$ operations (we ignore edge effects and other details) for *each* of the d^L product basis configurations, and the overall operation count is $O(2LD^3 d^L)$ which is *exponentially large* in system size even if we truncate MPS matrices. This would conclude the story, but there is an exponential speedup: contractions are carried out moving through the network from the left to the right, adding one matrix after another into the contracted part. The operational count is now $O((2L-1)D^3 d)$, which is linear in L and only weakly polynomial in

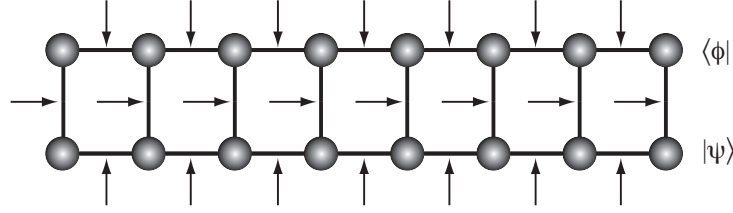


Fig. 10: Contraction scheme for the overlap of two MPS. All arrows point to contraction points.

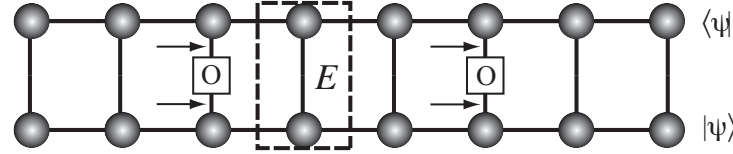


Fig. 11: Contraction scheme for a two-point correlator, indicating the transfer operator E .

D . This can best be seen by rearranging the overlap equation as

$$\begin{aligned}
 \langle \phi | \psi \rangle &= \sum_{\{\sigma\}} \tilde{M}^{\sigma_1*} \dots \tilde{M}^{\sigma_L*} M^{\sigma_1} \dots M^{\sigma_L} \\
 &= \sum_{\{\sigma\}} \tilde{M}^{\sigma_L\dagger} \dots \tilde{M}^{\sigma_1\dagger} M^{\sigma_1} \dots M^{\sigma_L} \\
 &= \sum_{\sigma_L} \tilde{M}^{\sigma_L\dagger} \left(\dots \left(\sum_{\sigma_2} \tilde{M}^{\sigma_2\dagger} \left(\sum_{\sigma_1} \tilde{M}^{\sigma_1\dagger} M^{\sigma_1} \right) M^{\sigma_2} \right) \dots \right) M^{\sigma_L}.
 \end{aligned}$$

If one works out the contractions from inside out, the first bracket costs $O(dD^3)$, ignoring that the first and last matrices are in reality vectors. The result is again a matrix, so the next bracket is essentially a product of three matrices (carried out as two matrix-matrix multiplications) and the cost is $O(2dD^3)$. Overall, we obtain the contraction count announced. This is the optimal scheme, and the contraction is exact; I am mentioning this because in the higher-dimensional generalizations of MPS contractions are not exact unless exponential complexity is accepted (which is of course impossible) and finding the optimal contraction scheme is NP hard (i.e. impossible to determine in practice).

Evaluating expectation values of operators is now very easy: assume we want to evaluate the expectation value of \hat{O} acting on site i , we just have to replace \sum_{σ_i} by $\sum_{\sigma_i, \sigma'_i} O_{\sigma_i, \sigma'_i}$, a double sum over the matrix elements of \hat{O} in the local basis, information which is trivially available. So the computational cost is hardly growing at all, and graphically, one would replace the vertical line at site i by the operator. If we consider the object

$$E^{(a_{\ell-1}a'_{\ell-1}), (a_{\ell}, a'_{\ell})} := \sum_{\sigma_{\ell}} A_{a_{\ell-1}, a_{\ell}}^{\sigma_{\ell}*} A_{a'_{\ell-1}, a'_{\ell}}^{\sigma_{\ell}}, \quad (49)$$

which is shown in Fig. 11, one can work out analytically or see graphically that it acts as a transfer matrix in the calculation of correlators, determining the decay of correlators by its eigenvalues. As is typical for transfer matrix calculations, all correlators are either long-ranged

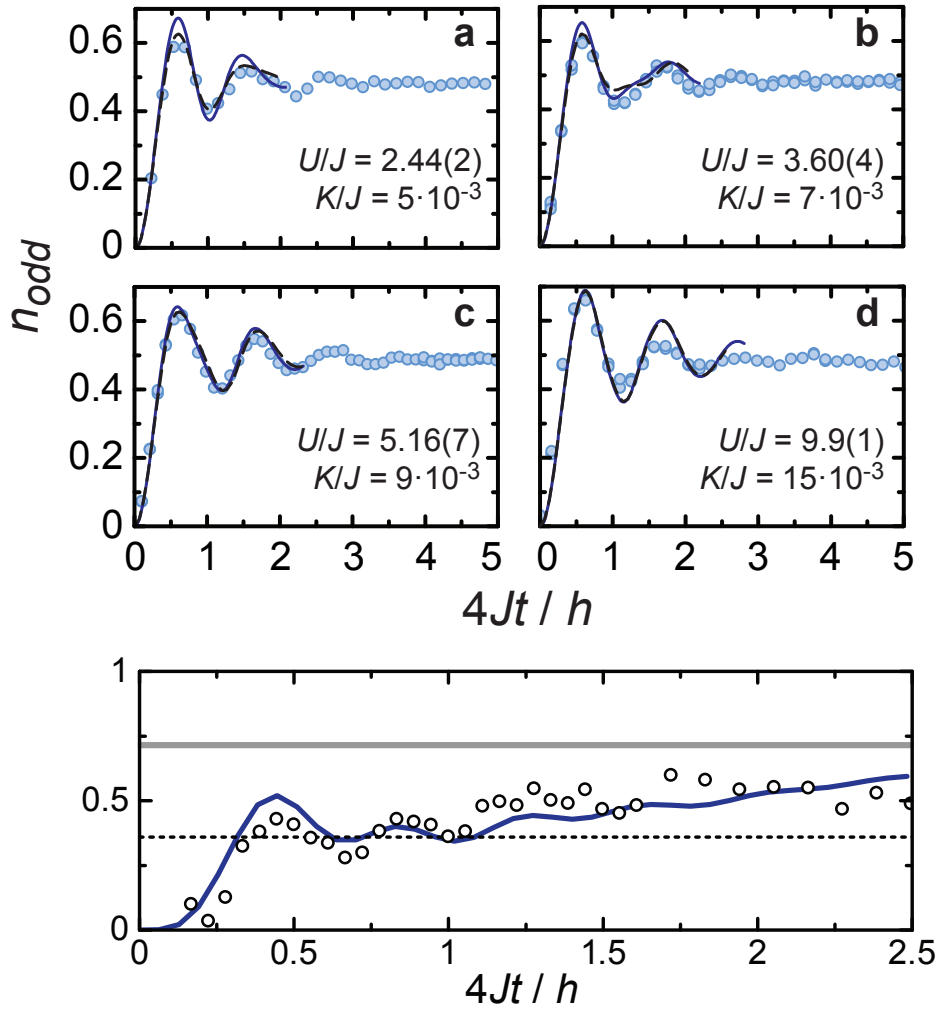


Fig. 12: Comparison between *t*DMRG simulations (lines) and experiment (points). In both cases, the starting wave function is a density wave in an optical lattice with one boson on even and no bosons on odd sites, that evolves according to a Bose-Hubbard-Hamiltonian with hopping amplitude J and on-site repulsion U . The upper panel shows how the density relaxes from 0 to 0.5 for various interaction strengths. For $U/J = 9.9$ one-dimensionality is partially lost, leading to disagreement. The lower panel shows $4\text{Re}\langle b_i^\dagger b_{i+1} \rangle$ as a function of time: nearest-neighbour quantum correlations are built up over time by the relaxation of the density wave. Taken from [32].

or decay as a superposition of exponentials. Critical power-law decays are not possible for MPS. Therefore, for a quantum state that has critical correlations, the MPS will approximate the power law decay by a superposition of exponentials, but eventually switch to an exponential decay on very long length scales, when only the slowest exponential decay in the superposition survives. Increasing D , the MPS will model the critical decay on increasing length scales.

As an example of the accuracy achieved nowadays both by experiment and theory for non-trivial strongly interacting systems, using time-dependent DMRG, consider the comparisons shown in Fig. 12 between *t*DMRG and a state-of-the-art quantum simulator using ultracold atoms in an optical lattice [32].

7 Finite-temperature simulations

So far, we have been considering pure-state calculations only; quantum mechanics generally deals with mixed states that are represented by (reduced) density operators $\hat{\rho}$, which are Hermitian, non-negative, and normalized (trace 1). In the early days of MPS-formulations of DMRG, it seemed advantageous to find a MPS-type *state* representation of $\hat{\rho}$ in order to “re-cycle” existing methods. This is indeed possible, using the concept of purification, first used in this context by [8], which is the Schmidt decomposition read backwards: given a density operator living on physical space P expressed in eigen-representation

$$\hat{\rho}_P = \sum_n \rho_n |n\rangle_P \langle n|, \quad (50)$$

we can interpret it as the reduced density operator on P for a state $|\psi\rangle_{PQ}$ living on a larger space PQ , where Q must be at least as large as P (in practice, simply a copy of P : instead of a spin chain, we simulate a spin ladder, etc.), and where the Schmidt decomposed form of $|\psi\rangle_{PQ}$ is

$$|\psi\rangle_{PQ} = \sum_n \sqrt{\rho_n} |n\rangle_P |n\rangle_Q. \quad (51)$$

The choice of the $|n\rangle_Q$ living on the auxiliary space Q is free as long as they form an orthonormal set; this gives a gauge degree of freedom to $|\psi\rangle_{PQ}$, the purification of $\hat{\rho}_P$

$$\hat{\rho}_P = \text{tr}_Q |\psi\rangle_{PQ} \langle \psi|_{PQ}. \quad (52)$$

Provided we know the purification, all further calculations can be done in a pure state framework on PQ . Expectation values are (normalization of $\hat{\rho}$ and $|\psi\rangle$ imply each other)

$$\langle \hat{O}_P \rangle_{\hat{\rho}_P} = \text{tr}_P \hat{O}_P \hat{\rho}_P = \text{tr}_P \hat{O}_P \text{tr}_Q |\psi\rangle_{PQ} \langle \psi|_{PQ} = \text{tr}_{PQ} \hat{O}_P |\psi\rangle_{PQ} \langle \psi|_{PQ} = \text{tr}_{PQ} \langle \psi| \hat{O}_P |\psi\rangle_{PQ}, \quad (53)$$

where we have used the cyclicity of the trace. Time evolution becomes

$$\hat{\rho}_P(t) = e^{-i\hat{H}t} \hat{\rho}_P e^{+i\hat{H}t} = e^{-i\hat{H}t} \text{tr}_Q |\psi\rangle_{PQ} \langle \psi|_{PQ} e^{+i\hat{H}t} = \text{tr}_Q |\psi(t)\rangle_{PQ} \langle \psi(t)| \quad (54)$$

with $|\psi(t)\rangle_{PQ} = e^{-i\hat{H}t} |\psi\rangle_{PQ}$. Hence we simply have to carry out a standard time-evolution of a pure state. The problem is, of course, that usually we do not know the eigenrepresentation of $\hat{\rho}_P$. But for the most important application, the thermal state $e^{-\beta\hat{H}}$, it can be calculated easily: $e^{-\beta\hat{H}} = e^{-\beta\hat{H}/2} \cdot \hat{I}_P \cdot e^{-\beta\hat{H}/2} = \text{tr}_Q e^{-\beta\hat{H}/2} |\rho_0\rangle_{PQ} \langle \rho_0| e^{-\beta\hat{H}/2}$, where $|\rho_0\rangle_{PQ}$ is the purification of $\hat{\rho}_P(\beta = 0) = \hat{I}_P$ (up to an irrelevant normalization). The infinite temperature ($\beta = 0$) reduced density operator factorizes between individual sites, and so does the purified state between pairs of sites in P and Q associated with each other. But the purification of the totally mixed state $\hat{\rho}_P(\beta = 0)$ on a site is nothing but a maximally entangled state between a site in P and the corresponding site in Q , for example the singlet state if the physical site contains a spin- $\frac{1}{2}$. Now this state can be encoded easily as an MPS on one pair of sites, and as the MPS factorizes between pairs of sites, the matrices have dimension 1 and can be written down by

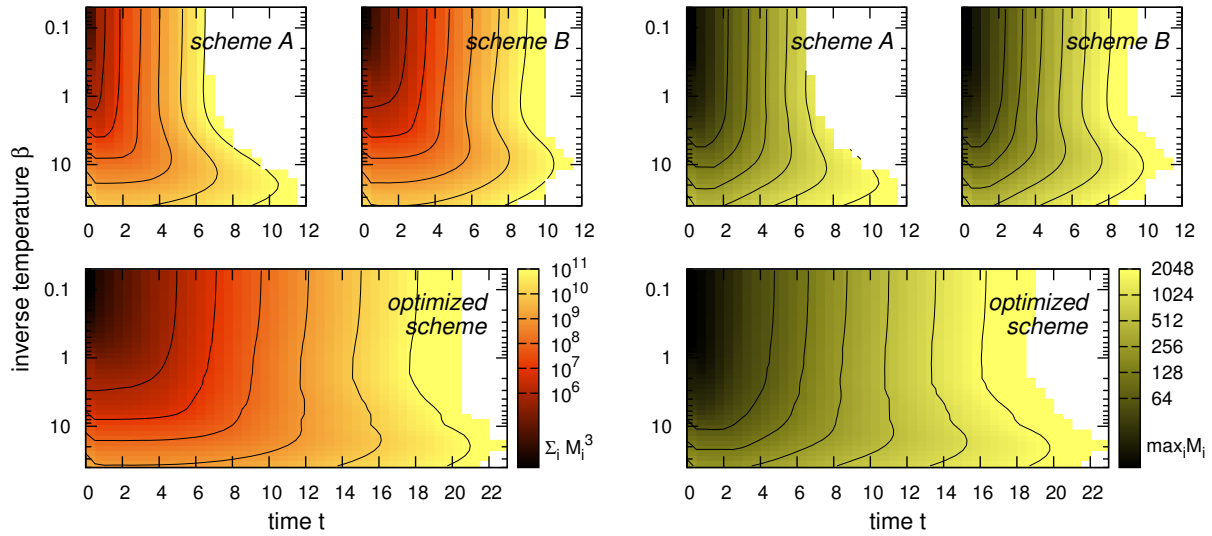


Fig. 13: Reachable times for the isotropic Heisenberg model with $\hat{B}^\dagger = \hat{A} = \hat{S}^\pm$ at the chain center: Schemes A, B are the conventional scheme [8] and the improved scheme [34] compared to the new scheme of [36]. The left panels show computational cost (proportional to the sum of the third power of the matrix dimensions), the right panels the maximal matrix dimension. Similar greyscales correspond to similar usage of resources. Adapted from [36].

hand. $|\rho_\beta\rangle_{PQ}$, a purification of $\hat{\rho}_P(\beta)$, is then obtained by an imaginary time evolution up to $\beta/2$. For time-dependent results obtained in this way and enhanced by a prediction technique, consider [33]. However, it was observed that times reached are relatively short.

Let us consider this in more detail. If we subject a thermal state to a real-time evolution using the same Hamiltonian, it remains unchanged as $[\hat{\rho}_P(\beta), \hat{H}] = 0$. In the purification approach, however, resource usage grows, because we are time-evolving the purified state which is not an eigenstate of \hat{H} , and the above reasoning for entanglement growth applies. We are therefore moving through increasing costly purifications of the same density operator. As there is a gauge freedom in the auxiliary state space, it might be possible to counteract this resource growth by a suitable transformation of auxiliary states. Indeed, it was found that evolving the auxiliary system backward in time, using $e^{+i\hat{H}t}$, resources will not grow [34], and this increase in efficiency also pays off for time-evolutions once a local operator has been applied. Substantial further improvement [35, 36] (and a better understanding of the procedure of [34]) is possible by exploiting the isomorphism between local bounded operators $\hat{B} : \mathcal{H} \mapsto \mathcal{H}$ and states $|\psi\rangle$ on $\mathcal{H} \otimes \mathcal{H}$. Our purified (MPS) state therefore has an (MPO) operator interpretation on \mathcal{H} , and we can discard the notion of purification entirely. In fact, in my view, this is the formulation to use in the future, the old purification approach having its usefulness because it allows recycling of tDMRG, whereas in the new approach, MPO manipulation routines are needed. The currently best approach found there is given as

$$\langle \hat{B}(2t)\hat{A} \rangle_\beta = Z(\beta)^{-1} \text{tr} \left([e^{i\hat{H}t} e^{-\beta\hat{H}/2} \hat{B} e^{-i\hat{H}t}] [e^{-i\hat{H}t} \hat{A} e^{-\beta\hat{H}/2} e^{i\hat{H}t}] \right), \quad (55)$$

where the objects in $[\cdot, \cdot]$ are treated as MPOs that are constructed by Trotterization and compression just as in normal time-evolution. An auxiliary space Q is not needed at all.

8 Ground states with MPS: DMRG

We have already seen that an imaginary time evolution is capable of yielding a ground state. The classic technique, which is more efficient, is by a variational minimisation within MPS space, and is what is traditionally referred to as DMRG, which is a variational minimisation technique in MPS space and hence sometimes, but rarely, also referred to as VMPS. For a given allowed maximum MPS matrix dimension D , the best MPS approximation to the ground state energy (and the ground state) is given by

$$\min \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \Leftrightarrow \min \left(\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle \right), \quad (56)$$

where we have introduced a Lagrangian multiplier λ to enforce normalization $\langle \psi | \psi \rangle = 1$ that will give the variational approximation to the ground state energy E_0 from above (because we are in a restricted search space). If we increase D , search space is enlarged (as it contains all MPS of smaller dimension), and the approximate energy will decrease monotonically, allowing for an extrapolation in $D \rightarrow \infty$, which is the exact limit.

Assuming we know the MPO representation of \hat{H} , which I postpone for a moment, we have to minimise the network represented in Fig. 14. This is a multilinear problem, as states $|\psi\rangle$ are products of the unknown M^σ , for which no immediately efficient strategy is known. A time-proven approach to the problem is the *alternating least square (ALS) method*, which runs as follows:

1. Start with a guess for the MPS extremizing energy, $\{M^{\sigma_i}\}$.
2. Pick a site i , $1 \leq i \leq L$, and consider all matrices M^{σ_j} , $j \neq i$, fixed and retain only M^{σ_i} as variables. Eq. (56) is then *quadratic* in M^{σ_i} , and minimization becomes a linear problem, leading to new M^{σ_i} minimizing energy within the “framework” provided by the other matrices.
3. Now pick another site as i , repeating step 2, until all sites have been visited often enough that λ does not decrease anymore. The resulting MPS is the variationally best approximation to the ground state, as is λ to the ground state energy.

Let us elaborate on these steps, because an efficient implementation strongly rests on details here. The starting guess can be random or be constructed by an iterative growth of a chain of length $2 \rightarrow 4 \rightarrow \dots \rightarrow L$, where at each step the chain is grown by inserting 2 sites at

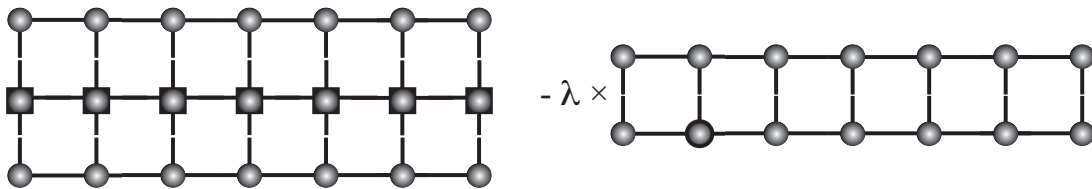


Fig. 14: Network to be contracted to extremize $\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle$.

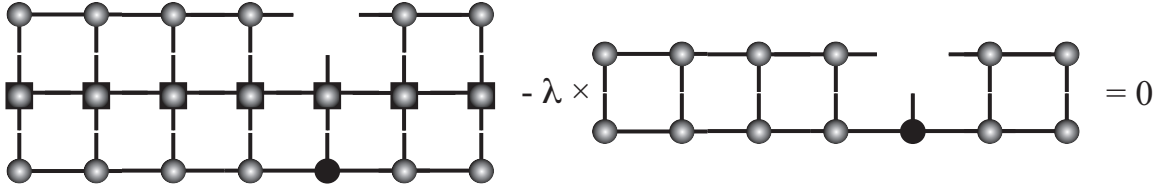


Fig. 15: Network to be contracted to extremize $\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle$. The unknown (variable) matrix M^{σ_i} is shown in black.

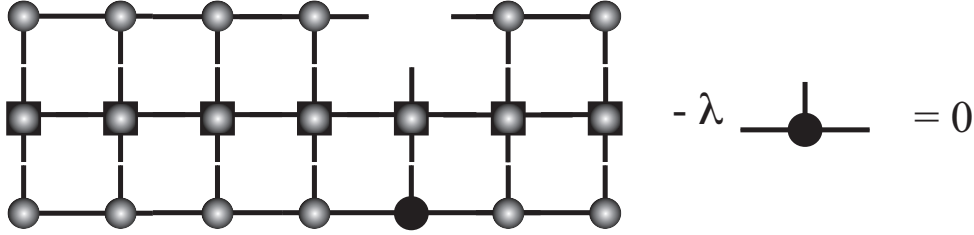


Fig. 16: Eigenvalue problem to be solved for extremizing $\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle$ provided the state $|\psi\rangle$ is in adequate mixed canonical form. Again, the unknown (variable) matrix M^{σ_i} is shown in black.

the center, and determining the two sets of M^σ for these 2 sites such that energy is minimized while keeping all previously found matrices fixed. This relates to the block plus site growth strategy exposed earlier, and is what traditionally is referred to as “infinite system DMRG”. Historically, this was even considered the core of the method, whereas it is rather the warm-up to the “real” algorithm! This statement has to be qualified to the extent that for translationally invariant systems the “warm-up” can be turned into a ground state search algorithm in its own right under the name of iDMRG [12, 37].

For carrying out the minimization step, we take the derivative

$$\frac{\partial}{\partial M^{\sigma_i}} \left(\langle \psi | \hat{H} | \psi \rangle - \lambda \langle \psi | \psi \rangle \right) \stackrel{!}{=} 0. \quad (57)$$

Graphically, this corresponds to Fig. 15, where the matrix with respect to which we take the derivative is simply removed, because it contributed linearly to the original network. If we now contract the two remaining networks, we have three free legs each and three legs connecting to the unknown M^{σ_i} . We can rewrite this as

$$\sum_{\sigma'_i a'_{i-1} a'_i} H_{\sigma_i a_{i-1} a_i, \sigma'_i a'_{i-1} a'_i} M_{\sigma'_i a'_{i-1} a'_i} = \sum_{\sigma'_i a'_{i-1} a'_i} N_{a_{i-1} a_i, a'_{i-1} a'_i} \delta_{\sigma_i, \sigma'_i} M_{\sigma'_i a'_{i-1} a'_i} \quad (58)$$

$$\equiv \sum_{\sigma'_i a'_{i-1} a'_i} N_{\sigma_i a_{i-1} a_i, \sigma'_i a'_{i-1} a'_i} M_{\sigma'_i a'_{i-1} a'_i}, \quad (59)$$

where the primed variables run over the legs connecting to M^{σ_i} and the unprimed variables run over the free legs. H and N are the contracted networks, and for convenience I have brought all indices down. The resulting equation has the form of a generalized eigenvalue problem,

$$H \mathbf{m} = \lambda N \mathbf{m}, \quad (60)$$

where the matrices are $(dD^2 \times dD^2)$ dimensional and the vector \mathbf{m} (from the reshaped matrices M^{σ_i}) is dD^2 -dimensional. As we are looking for the ground state, we have to find the lowest eigenvalue λ . With $D \sim 1000$ in practice, this cannot be found by full diagonalization techniques, but we can use *large sparse eigensolvers* such as provided by the Lanczos method (for a first introduction, see [22]).

Generalized eigenvalue problems can be potentially dangerous, if the condition number of matrix N becomes large. But if we assume that all matrices in $|\psi\rangle$ to the left of M^{σ_i} are left-normalized, $AAAA\dots$, and all matrices to the right of it are right-normalized, $\dots BBBB$, the normalization conditions imply that the entire network on the right collapses and only M^{σ_i} remains. Then $N = I$, and a simple eigenvalue problem $H\mathbf{m} = \lambda\mathbf{m}$ remains. In order to achieve this convenient situation, we simply take our starting state, bring it by (partial) normalization into form $AAAAAAMB BBBB$ (wherever we want to start) and replace random choices of locations i by systematic “sweeping” through the chain from right to left to right to left and so on (or inversely), carrying out one suitable normalization step at each iteration moving left or right to keep the seam between the left- and right-normalized matrices moving along with i . This sweeping procedure is exactly the finite-system DMRG invented by White in 1992 and the concurrent normalization procedure is nothing but White’s prediction algorithm [38].

A few more remarks are in order (for details, see [12]): contracting the left network (building H) might in principle be very costly, but in practice one keeps three separate parts, namely the part of H to the left of M^{σ_i} , the one on top (the part of the MPO), and the part of H to the right of M^{σ_i} . If we sweep left and right, these parts can be recycled or updated iteratively from previous steps, drastically reducing the numerical cost. At the same time, original DMRG considered two sites for optimization at once, i.e. pairs $M^{\sigma_i} M^{\sigma_{i+1}}$. Looking more closely at how this is done, one realizes that this slightly breaks the variational nature of the method and is a bit more costly numerically (by a factor of d); on the other hand, it is less plagued by the problem of our “single-site” method that it may sometimes get stuck in a non-global minimum (being non-variational can take you out of such dead ends). But this issue can be resolved elegantly by adding suitable “noise” to the procedure [12, 39].

How can we control accuracy? Extrapolating in $D \rightarrow \infty$ is fine in principle (energies go down monotonically, due to the variational nature of the procedure, other operators also converge monotonically as a rule of thumb), but if MPS make sense it is because the Schmidt coefficients decay rapidly, usually exponentially fast. But then we expect some kind of exponential convergence in D which however is not so clean as to allow an easy extrapolation. It is easier to monitor, as a function of D , the variation of the energy,

$$\langle \psi | \hat{H}^2 | \psi \rangle - (\langle \psi | \hat{H} | \psi \rangle)^2, \quad (61)$$

which can be evaluated without further approximation as the Hamiltonian MPO is exact, and to extrapolate quantities in the variation, which is 0 for $D \rightarrow \infty$ ($|\psi\rangle$ will be an exact eigenstate).

9 Constructing the MPO representation of a Hamiltonian

So far we have dodged the issue of constructing the MPO corresponding to a Hamiltonian. In fact, for short-ranged Hamiltonians this is quite easy, and as the construction is best understood by looking at an example, I will consider the Heisenberg model

$$\hat{H} = J \sum_{i=1}^{L-1} \left(\frac{1}{2} (\hat{S}_i^+ \hat{S}_{i+1}^- + \hat{S}_i^- \hat{S}_{i+1}^+) + \hat{S}_i^z \hat{S}_{i+1}^z \right) + h \sum_{i=1}^L \hat{S}_i^z. \quad (62)$$

This consists of operator strings of the type $\hat{I} \otimes \hat{I} \otimes \hat{S}^+ \otimes \hat{S}^- \otimes \hat{I} \otimes \hat{I} \dots$. To simplify the notation, I will introduce *operator valued* matrices in MPOs, namely

$$\hat{M}^{[i]} = \sum_{\sigma_i, \sigma'_i} M^{\sigma_i, \sigma'_i} |\sigma_i\rangle \langle \sigma'_i|. \quad (63)$$

Then the Hamiltonian will take the form $\hat{H} = \hat{M}^{[1]} \hat{M}^{[2]} \dots \hat{M}^{[L]}$. Let us imagine the construction of the Hamiltonian as the action of an automaton which has internal states (not to be confused with quantum states of our system). It starts from the right end of the chain in some internal state and moves through it to the left end. The action of the automaton is shown in Fig. 17. The automaton starts in internal state 1, acts, and once it has passed site 1, it should be in some final internal state and have produced exactly all terms that contribute to the Hamiltonian. This can be achieved as follows: we associate internal state 1 with “no non-trivial operator to the right”. Being in state 1, the automaton has five options at a site: adding another \hat{I} to the operator string (staying in 1), adding a \hat{S}^+ term (moving to state 2), a \hat{S}^z term (moving to state 3), or a \hat{S}^- term (moving to state 4). In any of those cases, the automaton now must add the term that completes the interaction at the next site, this is a $(J/2)\hat{S}^-$ term, a $J\hat{S}^z$ term, or a $(J/2)\hat{S}^+$ term. In any case, it moves into state 5, which corresponds to “completed interaction to the right”. At the same time, it can also move directly from state 1 to 5 by introducing a $h\hat{S}^z$ field term.

The action of the automaton can now be represented in matrix form where the row and column indices correspond to the outgoing and ingoing internal state of the automaton at each step. Taking into account that at site 1 the automaton must end in state 5 and that at site L it must start in state 1, the matrices read

$$\hat{M}^{[i]} = \begin{bmatrix} \hat{I} & 0 & 0 & 0 & 0 \\ \hat{S}^+ & 0 & 0 & 0 & 0 \\ \hat{S}^z & 0 & 0 & 0 & 0 \\ \hat{S}^- & 0 & 0 & 0 & 0 \\ h\hat{S}^z & (J/2)\hat{S}^- & J\hat{S}^z & (J/2)\hat{S}^+ & \hat{I} \end{bmatrix} \quad (64)$$

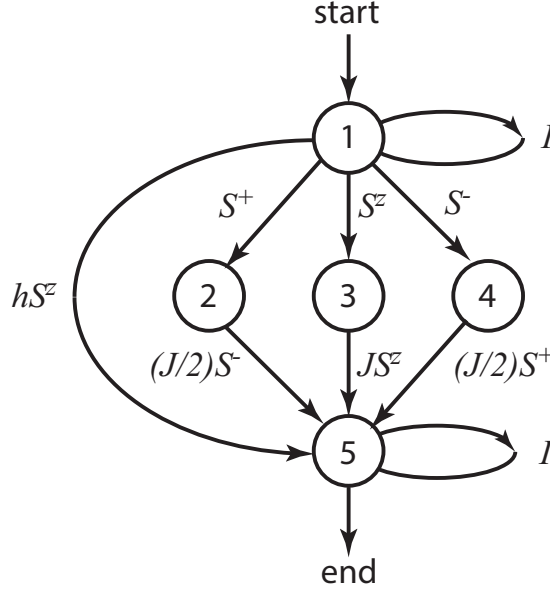


Fig. 17: States of the automaton constructing the Hamiltonian MPO for the Heisenberg model.

and on the first and last sites

$$\hat{M}^{[1]} = \begin{bmatrix} h\hat{S}^z & (J/2)\hat{S}^- & J^z\hat{S}^z & (J/2)\hat{S}^+ & \hat{I} \end{bmatrix} \quad \hat{M}^{[L]} = \begin{bmatrix} \hat{I} \\ \hat{S}^+ \\ \hat{S}^z \\ \hat{S}^- \\ h\hat{S}^z \end{bmatrix}. \quad (65)$$

We can identify $\tilde{D} = 5$ as the dimension of the Hamilton MPO which is given by the above matrices. Similarly, other Hamiltonians can be constructed; for longer-ranged interactions, we have to introduce further internal states such that the automaton can keep track of when to complete an interaction. For a more elaborate discussion, see [12] and references therein.

10 Dynamical DMRG

Dynamical DMRG is the denomination for methods for the calculation of frequency-dependent Green's functions and spectral functions, hence the name which is not to be confused with *time-dependent* DMRG for the calculation of out-of-equilibrium dynamics.

The fundamental object under study are objects such as

$$S_\eta(\omega) := \left\langle 0 \left| \hat{O}^\dagger \frac{1}{E_0 + \omega + i\eta - \hat{H}} \hat{O} \right| 0 \right\rangle, \quad (66)$$

where $\eta = 0^+$ (numerically, it will be a small, but finite positive number) and $|0\rangle$ is the ground state of \hat{H} . This is nothing but the Fourier transform (with a numerical convergence factor $e^{-\eta t}$) of the $T = 0$ Green's function

$$G_O(t) = \langle 0 | \hat{O}^\dagger(t) \hat{O}(0) | 0 \rangle \quad (t > 0) \quad (67)$$

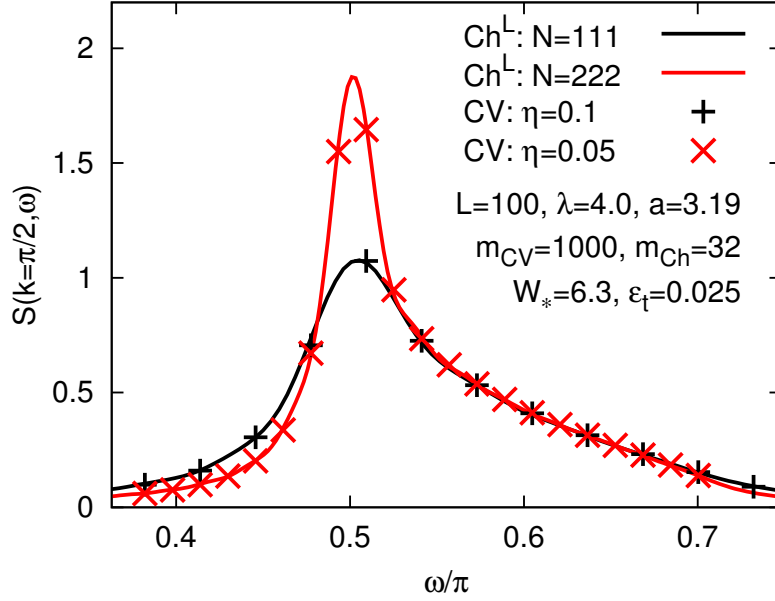


Fig. 18: Comparison between correction vector (CV) and Chebyshev vector (Ch) results for the $k = \frac{\pi}{2}$ spectral function of the isotropic Heisenberg antiferromagnet. Agreement is excellent, but the Chebyshev results are obtained orders of magnitude faster. From [45].

where the Heisenberg picture is assumed. For available $|0\rangle$, various techniques have been proposed which are not restricted to MPS, but can be efficiently implemented there.

The fastest (but often least precise) technique is the *continuous fraction* or Lanczos approach. This approach was pioneered in [40], first used in the classic DMRG context by [41], but can be made much more precise if expressed in MPS language [42]. Starting from $|q_1\rangle = \hat{O}|0\rangle / \|\hat{O}|0\rangle\|$, it generates a sequence of orthonormal Krylov vectors (also known as Lanczos vectors) $|q_m\rangle$ as

$$\beta_m |q_{m+1}\rangle = \hat{H}|q_m\rangle - \alpha_m |q_m\rangle - \beta_{m-1} |q_{m-1}\rangle, \quad (68)$$

where the α_m and β_m (calculated by normalizing $|q_{m+1}\rangle$) are the diagonal and off-diagonal elements of a tridiagonal matrix representation of \hat{H} : $\alpha_m = \langle q_m | \hat{H} | q_m \rangle$ and $\beta_m = \langle q_{m+1} | \hat{H} | q_m \rangle$. Then it can be shown by matrix inversion that

$$S_\eta(\omega) = \frac{\langle 0 | \hat{O}^\dagger \hat{O} | 0 \rangle}{E + i\eta - \alpha_1 - \frac{\beta_1^2}{E + i\eta - \alpha_2 - \frac{\beta_2^2}{E + i\eta - \alpha_3 - \dots}}}. \quad (69)$$

The attractive feature of this procedure is that the generation of Krylov vectors is usually available for free as the necessary algorithm already is part of the Lanczos large sparse matrix diagonalization routine needed in the ground state search.

The current gold standard, but also slowest, approach is the *correction vector* approach [43,44]. One introduces the so-called correction vector

$$|c\rangle = \frac{1}{E_0 + \omega + i\eta - \hat{H}} \hat{O} |0\rangle, \quad (70)$$

such that the Green's function reads $G_\eta(\omega) = \langle 0 | \hat{O}^\dagger | c \rangle$. $|c\rangle$ is determined by solving the large sparse equation system

$$(E_0 + \omega + i\eta - \hat{H})|c\rangle = \hat{O}|0\rangle, \quad (71)$$

for which reliable, but quite costly techniques such as GMRES are available [22]. Note that it is not a good idea to “square” the system as originally proposed in order to make it hermitian, because the new condition number is now the square of the old one, drastically slowing down convergence. Taking the limit $\eta \rightarrow 0$ also implies strongly increasing D of the correction vector MPS to maintain the desired accuracy, for reasons not fully understood at the moment.

More recently, a new technique was proposed which expands the Green's function in terms of *Chebyshev polynomials* [45]; I refer to the literature as it is not yet a widely used standard method; but from Fig. 18 it should be clear that it is a serious contender for the correction vector method, because it achieves similar accuracy orders of magnitude faster. – It remains, however, to be seen whether this advantageous scenario continues to hold for more complex systems.

Concluding this section, let me mention that both at $T = 0$ and $T > 0$ (where the techniques just discussed do not apply) frequency-dependent can be successfully obtained by a combination of time-dependent DMRG for the calculation of two-time correlators and subsequent Fourier transformation (for $T = 0$, see [6], for $T > 0$ see [33]). Frequency resolution is limited by the finite range in time of tDMRG which necessitates some exponential damping of the real-time data.

11 Outlook: DMRG in two dimensions

MPS and DMRG are obviously best suited to the study of one-dimensional quantum systems. However, there has been enormous interest in the notoriously elusive physics of two-dimensional quantum systems for decades, ranging from frustrated magnets to high- T_c superconductors, with a recent surge of interest in the physics of topological quantum spin liquids. Therefore it is not surprising that there have been numerous attempts to apply DMRG (MPS) to

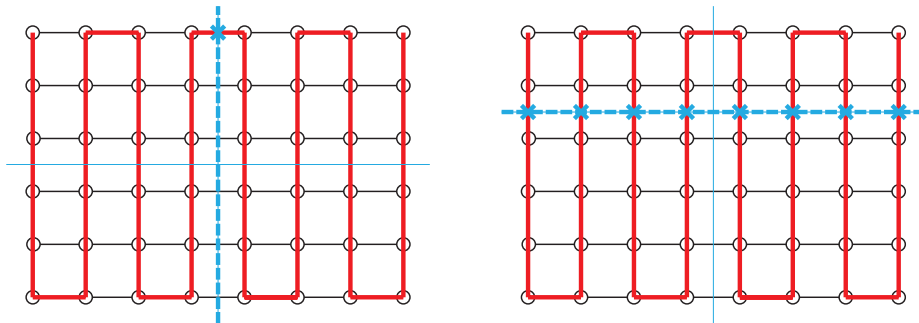


Fig. 19: Two-dimensional DMRG setup: The 2D lattice is explored by a one-dimensional snake, for which one formulates an MPS. Horizontal interactions become long-ranged, but the real problem is that encoding the entanglement of vertically separated bipartitions of the system becomes exponentially expensive.

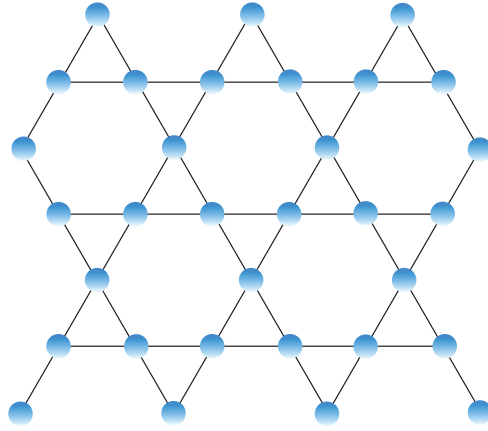


Fig. 20: The kagome lattice, whose name derives from the structure of Japanese bamboo baskets, but can be found in natural magnetic substances.

two-dimensional quantum systems, seriously starting with [38] and mainly focusing on Heisenberg and t - J -models in two dimensions.

The fundamental idea is to map (see Fig. 19) the two-dimensional lattice to a one-dimensional snake winding through the system, for which an MPS can be formulated. Technically, the price to pay is that short-ranged interactions perpendicular to the snake direction can become long-ranged (consider horizontal interactions in Fig. 19). In the beginning, this was considered to be the main reason for the comparatively disappointing results obtained (the reasoning need not concern us here). But in reality, the reason is given by the fact that entanglement for ground states of short-ranged Hamiltonians scales linearly with system size (area law [46]), i.e. $S \propto L$. As we have seen, a reduced density operator of dimension D can at most encode entanglement $S = \ln D$. Now the reduced density operators generated by an MPS of dimension D have at most dimension D themselves; hence we need MPS of (at least, because the distribution of the weights enters also) dimension 2^D to encode entanglement S . If we consider the vertical cut through the system, it is crossed by the snake only once, and the matrices therefore have to be *exponentially large* in the vertical size of the lattice, which strongly limits achievable system sizes. On the upside, horizontal cuts are crossed by the snake L times, and we have D^L coefficients if we think about the bottom and top parts of the system being represented by single matrices, and we can encode entanglement $S = L \ln D$, matching the area law. The size restriction therefore only applies in one spatial direction, such that simulations have focused on long stripe systems $L_x \gg L_y$.

As tensorial generalizations of MPS such as PEPS [9], MERA [47] or iPEPS [48] do not suffer from exponential growth of resources in two dimensions, they were at first considered vastly superior to the essentially unsuitable DMRG approach in two dimensions. However, it was realized that they suffer not only from scaling polynomially much worse in tensor dimension (up to D^{16} as opposed to D^3 for MPS operations), which would always be preferable to exponential growth in resources for sufficiently large system sizes, but also from quite severe issues with normalizability and the conditioning of the arising linear algebra problems as well as further approximations involved in the contractions. Therefore, progress in that direction has not been

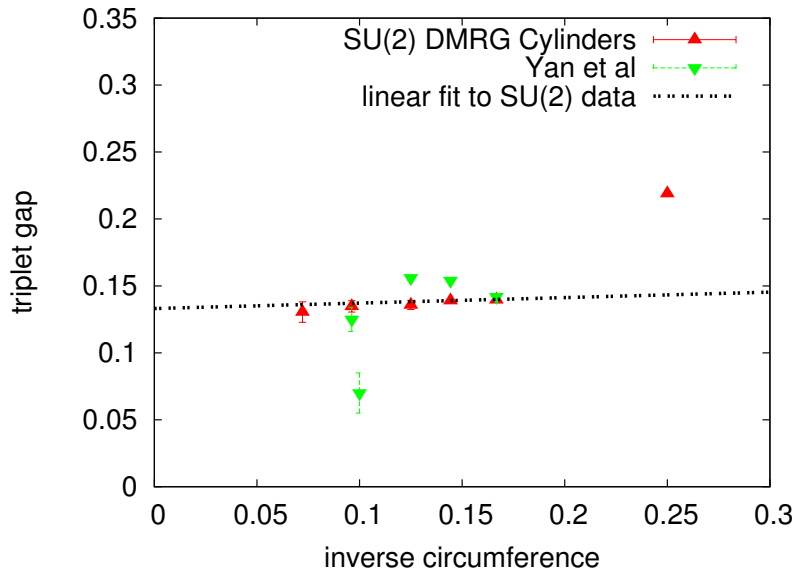


Fig. 21: Size of the (triplet) gap of the Heisenberg antiferromagnet on the kagome lattice versus inverse cylinder width obtained in two large-scale DMRG simulations. The more advanced $SU(2)$ symmetric simulations allow for the first time for a reliable extrapolation to the thermodynamic limit. Reproduced from [50].

quite as spectacular as originally hoped for, which has left at least for now quite a bit of space for brute-force, but numerically well controlled DMRG applications, which have progressed both due to more powerful computers and progress in highly efficient implementations of the algorithm.

To illustrate the state of the art, let me mention the recent large-scale numerical studies of the isotropic Heisenberg antiferromagnet on the kagome lattice illustrated in Fig. 20, which have stirred up a lot of interest [49, 50]. The nature of the ground state has been discussed since the late eighties, without any conclusive answer, but numerous competing proposals so far. The most likely candidate for the ground state (in the view of many, the definite solution) could be identified recently as a quantum spin liquid by [49], more precisely, using $D \sim 17000$ and exploiting non-Abelian symmetries of the Hamiltonian in the largest DMRG study done so far, as a Z_2 topological quantum spin liquid [50]; lattices studied were on cylinders of a width of up to 18 lattice spacings and a length of up to 70 lattice spacings. For an example, consider Fig. 21, where the first reliable extrapolation of the (triplet) gap of the model is shown.

References

- [1] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
- [2] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995)
- [3] J. Dukelsky, M.A. Martín-Delgado, T. Nishino and G. Sierra, Europhys. Lett. **43**, 457 (1998)
- [4] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004)
- [5] A.J. Daley, C. Kollath, U. Schollwöck and G. Vidal, J. Stat. Mech.: Theor. Exp. P04005 (2004)
- [6] S.R. White and A.E. Feiguin, Phys. Rev. Lett. **93**, 076401 (2004)
- [7] F. Verstraete, D. Porras and J.I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004)
- [8] F. Verstraete, J.J. Garcia-Ripoll and J.I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004)
- [9] F. Verstraete and J.I. Cirac, cond-mat/0407066 (2004)
- [10] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004)
- [11] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005)
- [12] U. Schollwöck, Ann. Phys. **326**, 96 (2011)
- [13] F. Verstraete, V. Murg and J.I. Cirac, Adv. Phys. **57**, 143 (2008)
- [14] A. Einstein, B Podolsky and N. Rosen, Phys. Rev. **47**, 777 (1935)
- [15] J. Bell, Physics **1**, 195 (1964)
- [16] M.A. Nielsen and I.L. Chuang: *Quantum computation and quantum information* (Cambridge University Press, 2000)
- [17] I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987)
- [18] I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Comm. Math. Phys. **115**, 477 (1988)
- [19] A. Klümper, A. Schadschneider and J. Zittartz, Europhys. Lett. **24**, 293 (1993)
- [20] A. Kolezhuk, R. Roth and U. Schollwöck, Phys. Rev. Lett. **77**, 5142 (1996)
- [21] A. Kolezhuk and H.J. Mikeska, Phys. Rev. Lett. **80**, 2709 (1998)
- [22] L.N. Trefethen and D. Bau III: *Numerical linear algebra* (SIAM, Philadelphia, 1997)
- [23] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)

- [24] R. Bulla, T.A. Costi and T. Pruschke, *Rev. Mod. Phys.* **80**, 395 (2008)
- [25] A. Weichselbaum, F. Verstraete, U. Schollwöck, J.I. Cirac and J. von Delft, *Phys. Rev. B* **80**, 165117 (2009)
- [26] I.P. McCulloch, *J. Stat. Mech.: Theor. Exp.* P10014 (2007)
- [27] M. Suzuki, *Prog. Theor. Phys.* **56**, 1454 (1976)
- [28] E.H. Lieb and D. W. Robinson, *Commun. Math. Phys.* **28**, 251 (1972)
- [29] C. Kollath, U. Schollwöck and W. Zwerger, *Phys. Rev. Lett.* **95**, 250402 (2006)
- [30] M. Cheneau, P. Barnettler, D. Poletti, M. Endres, P. Schauß, T. Fukuhara, C. Gross, I. Bloch, C. Kollath and S. Kuhr, *Nature* **481**, 484 (2012)
- [31] M. Cramer, A. Flesch, I.P. McCulloch, U. Schollwöck and J. Eisert, *Phys. Rev. Lett.* **101**, 063001 (2008)
- [32] S. Trotzky, Y.-A. Chen, A. Flesch, I.P. McCulloch, U. Schollwöck, J. Eisert and I. Bloch, *Nature Physics* **8**, 325 (2012)
- [33] T. Barthel, U. Schollwöck and S.R. White, *Phys. Rev. B* **79**, 245101 (2009)
- [34] C. Karrasch, J.H. Bardarson and J.E. Moore, *Phys. Rev. Lett.* **108**, 227206 (2012)
- [35] T. Barthel, U. Schollwöck and S. Sachdev, *arXiv:1212.3570* (2012)
- [36] T. Barthel, *arXiv:1301.2246* (2013)
- [37] I.P. McCulloch, *arXiv:0804.2509* (2008)
- [38] S.R. White, *Phys. Rev. Lett.* **77**, 3633 (1996)
- [39] S.R. White, *Phys. Rev. B* **72**, 180403 (2005)
- [40] E.R. Gagliano and C. A. Balseiro, *Phys. Rev. Lett.* **59**, 2999 (1987)
- [41] K. Hallberg, *Phys. Rev. B* **52**, 9827 (1995)
- [42] P.E. Dargel, A. Wöllert, A. Honecker, I.P. McCulloch, U. Schollwöck and T. Pruschke, *Phys. Rev. B* **85**, 205119 (2012)
- [43] T.D. Kühner and S.R. White, *Phys. Rev. B* **60**, 335 (1999)
- [44] E. Jeckelmann, *Phys. Rev. B* **66**, 045114 (2002)
- [45] A. Holzner, A. Weichselbaum, I.P. McCulloch, U. Schollwöck and J. von Delft, *Phys. Rev. B* **83**, 195115 (2011)

- [46] J. Eisert, M. Cramer and M.B. Plenio, Rev. Mod. Phys. **82**, 277 (2010)
- [47] G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007)
- [48] J. Jordan, R. Orús, G. Vidal, F. Verstraete and J.I. Cirac,
Phys. Rev. Lett. **101**, 250602 (2008)
- [49] S. Yan, D. A. Huse and S.R. White, Science **332**, 1173 (2011)
- [50] S. Depenbrock, I.P. McCulloch and U. Schollwöck, Phys. Rev. Lett. **109**, 067201 (2012)

17 Entanglement and Tensor Network States

Jens Eisert

Freie Universität Berlin

Dahlem Center for Complex Quantum Systems

Contents

1	Correlations and entanglement in quantum many-body systems	2
1.1	Quantum many-body systems	2
1.2	Clustering of correlations	4
1.3	Entanglement in ground states and area laws	5
1.4	The notion of the ‘physical corner of Hilbert space’	10
2	Matrix product states	11
2.1	Preliminaries	11
2.2	Definitions and preparations of matrix product states	13
2.3	Computation of expectation values and numerical techniques	17
2.4	Parent Hamiltonians, gauge freedom, geometry, and symmetries	23
2.5	Tools in quantum information theory and quantum state tomography	27
3	Higher-dimensional tensor network states	29
3.1	Higher-dimensional projected entangled pair states	29
3.2	Multi-scale entanglement renormalization	32
4	Fermionic and continuum models	34
4.1	Fermionic models	34
4.2	Continuum models	35

1 Correlations and entanglement in quantum many-body systems

1.1 Quantum many-body systems

In this chapter we will consider *quantum lattice systems* as they are ubiquitous in the condensed matter context or in situations that mimic condensed matter systems, as provided, say, by systems of cold atoms in optical lattices. What we mean by a quantum lattice system is that we think that we have an underlying lattice structure given: some lattice that can be captured by a graph. The vertices of this graph are associated with a quantum degree of freedom each, referred to as *constituents*, while edges correspond to neighbourhood relations. Interactions in the physical system are usually *local*, which means that all constituents only directly interact with finitely many neighbours on the lattice. Particularly important is the situation when all interactions except from direct nearest neighbour interactions can be safely neglected. Quantum lattice models of this type capture strongly correlated materials often exhibiting interesting electronic and magnetic properties. They serve as theoretical laboratories allowing to study features of topological order and non-conventional phase transitions. Quantum systems well modelled by lattice models in this sense also show a wealth of phenomenology in out-of-equilibrium situations, to mention only a few reasons why this kind of physical system is interesting.

In this chapter, we will provide a brief introduction into *tensor network approaches to the study of quantum lattice models*. The position taken may be slightly unusual in the sense that a rather strong emphasis is put onto methods and ideas of description, and not so much on the phenomenology itself (which can then be derived from such a description, needless to say). Given that it is the very development of the toolbox of tensor network methods itself that is being reviewed here, one that has led to many recent new insights, this seems a healthy standpoint.

But there is yet another shift of emphasis that may be somewhat unexpected: namely that rather quantum states and not so much Hamiltonians are in the focus of attention. Here it is mostly the very nature of ground and thermal states themselves that is being considered and studied, while Hamiltonians primarily reenter via the concept of a parent Hamiltonian. The main message of this book chapter can be summarised in a single paragraph:

Many natural quantum lattice models have ground states that are little, in fact very little, entangled in a precise sense. This shows that ‘nature is lurking in some small corner of Hilbert space’, one that can be essentially efficiently parametrized. This basic yet fundamental insight allows for a plethora of new methods for the numerical simulation of quantum lattice models using tensor network states, as well as a novel toolbox to analytically study such systems.¹

¹In this book chapter, we will discuss the elements of this framework, while at the same time we cannot provide a comprehensive review. This chapter will still contain slightly more material than what is covered in the course. For recent reviews covering related topics, see Refs. [1–5].

1.1.1 Quantum lattice models

We start by discussing the concept of a quantum lattice model. The underlying *graph* $G = (V, E)$ capturing the *lattice* may, in principle, be any graph, where V is the vertex and E the edge set. $\text{dist}(\cdot, \cdot)$ is then the graph-theoretical distance, so the minimum number of steps one has to walk on the graph in order to get from one vertex to another. Particularly important are, however, regular graphs, and even more so particularly common regular graphs. In fact, most of the time we will be concerned with simple *cubic lattices* $V = L^{\mathcal{D}}$ in dimension \mathcal{D} , and specifically one-dimensional lines for which $\mathcal{D} = 1$. But also other lattices such as *triangular lattices* or *Kagome lattices* often emerge naturally. $n = |V|$ is referred to as the *system size*. The quantum degree of freedom at each vertex can be a spin system of dimension d – which will be the situation in the focus of attention in this chapter – or a *bosonic* or a *fermionic* degree of freedom (which we touch upon). The entire Hilbert space of the system is hence given by $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$ in case of spin models. For bosons and fermions we consider the *Fock space*. Say, if \mathcal{K} is the Hilbert space associated with a single fermion, then we make use of the Fock space $\mathcal{F} = \wedge^*(\mathcal{K})$. Clearly, the dimension of the Hilbert space grows exponentially with the system size, specifically $\dim(\mathcal{H}) = \Omega(d^n)$ for a spin model, which means that a numerical exact diagonalisation of the underlying Hamiltonian is doomed to failure for already moderately large system sizes n . In fact, a naive diagonalisation of the Hamiltonian without exploiting any additional structure would require $O(d^{3n})$ operations, clearly infeasible for large quantum systems.

1.1.2 Local Hamiltonians

All Hamiltonians considered here will feature finite-ranged interactions, which means that they can be written in the form

$$H = \sum_{j \in V} h_j, \quad (1)$$

where each h_j is non-trivially supported only on finitely many sites in V (but not necessarily on site j only). It is called *k-local* if each h_j is supported on at most k sites, and *geometrically k-local* if each h_j is supported on a V_j with $\max_{a,b \in V_j} \text{dist}(a, b) = k - 1$. This is a most natural situation: Each constituent interacts then merely with its immediate neighbours in the lattice. We will restrict attention to such geometrically *k-local* Hamiltonians, in fact with few exceptions to nearest-neighbour models. One also often writes $\langle j, k \rangle$ for neighbours, so for sites $j, k \in V$ such that $\text{dist}(j, k) = 1$. Similarly, one also calls any observable that is non-trivially supported only on neighbouring sites a *local observable*.

There are a number of famous examples of such local Hamiltonians. A Hamiltonian that has specifically been studied countless many times – for good reasons – is the XY-model Hamiltonian on a one-dimensional line with n sites [6], where $V = \{1, \dots, n\}$ and $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$,

$$H = -\frac{1}{2} \sum_{\langle j, k \rangle} \left(\frac{1+\gamma}{4} X^{(j)} X^{(k)} + \frac{1-\gamma}{4} Y^{(j)} Y^{(k)} \right) - \frac{\lambda}{2} \sum_{j \in V} Z^{(j)}, \quad (2)$$

where $\gamma \in \mathbb{R}$ is called the *anisotropy parameter* and $\lambda \in \mathbb{R}$ is the *magnetic field*. The matrices

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (3)$$

are the familiar *Pauli matrices*, the index referring to the site they are supported on. It should be clear that this Hamiltonian is a local Hamiltonian of the above form. Simple as this model is, it can be easily analytically solved by considering it as a problem of free fermions. It already features notions of criticality, and is indeed often studied as a paradigmatic lattice model – not the least because everything imaginable can be said about it. The situation of $\gamma = 0$ is particularly important; then the model is also called *isotropic XY model* or *XX model*.

1.1.3 Boundary conditions

Depending on whether one identifies for the one-dimensional line $V = \{1, \dots, n\}$ the site $n+1$ with the site 1 or not, one says that one has *open* or *periodic* boundary conditions. We will also consider open and periodic boundary conditions for other cubic lattices, where $V \in L^D$. For periodic boundary conditions, one encounters then the topology of a torus in D dimensions.

1.1.4 Ground states, thermal states, and spectral gaps

The lowest energy eigenvectors of the Hamiltonian, so the normalized state vectors that minimise $\langle \psi | H | \psi \rangle$, form a Hilbert space \mathcal{G} , the *ground space*; one also often refers to the *ground state manifold*. If the ground space is one-dimensional, the ground state is *unique*, otherwise it is called *degenerate*. Ground states often capture the low temperature physics and their study is ubiquitous in theoretical physics. The energy associated with the ground space is the *ground state energy*, usually denoted as E_0 . *Ground state expectation values* will be denoted as $\langle O \rangle$, $O \in \mathcal{B}(\mathcal{H})$ being some observable. Particularly important are *local observables* O_A which are supported on finitely many sites $A \subset V$ only (actually most prominently on just a single site). The *Hamiltonian gap* is the energy gap from the ground space to the first excited state, so

$$\Delta E = \inf_{|\psi\rangle \in \mathcal{H} \setminus \mathcal{G}} \langle \psi | H | \psi \rangle - E_0. \quad (4)$$

If $\Delta E = 0$ for a family of Hamiltonians in the thermodynamic limit of $n \rightarrow \infty$, then one says that the system is *gapless* or *critical*. Such critical models can be beautifully captured in the framework of *conformal field theory* which is outside the scope of this book chapter. If a positive gap exists in the thermodynamic limit, it is *gapped*.

1.2 Clustering of correlations

Since direct interactions are local and constituents directly see their immediate neighbours only, one should expect that correlation functions between different constituents somehow decay with the distance in the lattice. The correlation functions hence should be expected to inherit the locality of interactions. It turns out that this is indeed the case.

1.2.1 Clustering of correlations in gapped models and correlation length

Specifically, for gapped models, correlation functions always decay exponentially with the distance. This effect is also called *clustering of correlations*. Nearby lattice sites will still be correlated to some extent, but these correlations become negligible for large distances. So if for a family of models $\Delta E > 0$ (and under very mild conditions on the lattice G which are always satisfied for natural finite-dimensional lattices), then [8]

$$|\langle O_A O_B \rangle - \langle O_A \rangle \langle O_B \rangle| \leq C e^{-\text{dist}(A,B) \Delta E / (2v)} \|O_A\| \|O_B\|, \quad (5)$$

for some suitable constant $C > 0$. The length scale

$$\xi := \frac{2v}{\Delta E} > 0 \quad (6)$$

emerging here is the *correlation length*: it is the characteristic length scale on which correlations disappear. The feature of clustering of correlations has long been suspected to be generically valid in gapped models and has been ‘known’ for a long time. A rigorous proof of this can be obtained in a quite beautiful way using *Lieb-Robinson bounds* [7], which are bounds to the speed of information propagation in time in quantum lattice models with local Hamiltonians, in terms of the *Lieb-Robinson velocity* $v > 0$. By using a clever representation in the complex plane, one can relate this statement – which as such relates to dynamical features – to ground state properties [8]. $\|\cdot\|$ in the above expression is the *operator norm*, so the largest singular value: It grasps the ‘strength’ of the observable. Again, if A and B are far away in gapped models, in that $\text{dist}(A, B)$ are much larger than the correlation length, then the correlation function will essentially disappear.

1.2.2 Algebraic decay in gapless models

For gapless models the above is no longer true. Generically, correlation functions of gapless models decay algebraically with the distance in the lattice. Then there is no longer a length scale associated with the decay of correlation functions. *Conformal field theory* provides a framework of systems exhibiting a conformal symmetry, which can be applied to the study of critical quantum lattice systems.

1.3 Entanglement in ground states and area laws

Yet, there is a stronger form of locality inherited by the locality of interactions than merely the decay of two-point correlation functions between any two sites. These stronger forms of locality are captured using concepts of entanglement in quantum many-body systems. Indeed, the insight that concepts and methods of *entanglement theory* – as they originally emerged in quantum information theory – can be applied to the study of quantum many-body systems triggered an explosion of interest in the community. The significance of this for grasping quantum many-body systems in condensed-matter physics with tensor networks will become clear in a minute.

1.3.1 Entanglement entropies

Let us imagine we have a gapped lattice model prepared in the ground state, based on some lattice $G = (V, E)$. We now single out a certain subset $A \subset V$ of sites, some region, and consider its complement $B := V \setminus A$. This subset will be associated with a reduced state $\rho_A = \text{tr}_B(\rho)$. The reduced state alone will allow to compute every expectation value of observables supported on A ; it is obtained by fixing an orthonormal basis in B and taking the partial trace. Now, what will the von-Neumann entropy

$$S(\rho_A) = -\text{tr}(\rho_A \log_2 \rho_A) \quad (7)$$

of the state ρ_A be? Of course, the entropy of the entire ground state ρ will vanish, so $S(\rho) = 0$, it being a pure state, but this is not true for the entropy of reduced states. If the ground state is unique, so if it is a pure state, which we are assuming here, this entropy reflects the *degree of entanglement* [9] of the system A with respect to its complement. If A and B are in a product state and no entanglement is present, then $S(\rho_A) = 0$. Otherwise, the entropy will be larger the more entangled the sub-systems are, being bounded from above by the maximum value $S(\rho_A) \leq |A| \log_2(d)$.

Quantum correlations make the entropy of reduced states become non-vanishing. In fact, according to a meaningful axiomatic quantification of asymptotic entanglement manipulation, the von-Neumann entropy *uniquely* quantifies the entanglement content [10] in a sense. This *entanglement measure* is called *entropy of entanglement* or *entanglement entropy*. Note that this is only a meaningful measure of entanglement, an entanglement monotone, as one says, for pure states; so in our context, if the ground state is non-degenerate. For mixed states, one can still compute the entropy of the reduced state, but this would no longer amount to an entanglement monotone (and this quantity then no longer captures the entanglement in the ground state, but has to be replaced by other measures that we will discuss below).

1.3.2 Area laws for the entanglement entropy

So how does $S(\rho_A)$ scale with the size of the $|A|$ of the region A ? Detailed knowledge about the entropy will here be less important than the general scaling behavior in the asymptotic limit of large regions. Questions of a similar type have a long tradition and were first asked in the context of the scaling of black hole entropies [11]. Naively, but at the same time naturally, one might expect this quantity to scale *extensively* with the size $|A|$: This would mean that $S(\rho_A) = O(|A|)$. After all, entropies of Gibbs states in statistical mechanics are known to scale extensively, so one might think that the same intuition may be applied here. And indeed, for ‘generic states’, so for random states, this is true with overwhelming probability. One can rigorously define random states using the Haar measure of the unitaries acting on \mathcal{H} , and finds that the expected entanglement entropy indeed takes the maximum value $|A| \log_2(d)$, up to a correction that is exponentially small in the size of B .

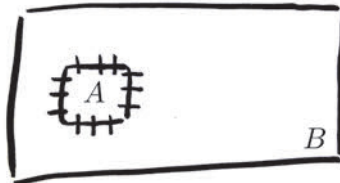
This intuition, however, is not correct. Instead, one finds that the entropy scales as the *boundary*

area of A , so

$$S(\rho_A) = \mathcal{O}(|\partial A|). \quad (8)$$

One then also says that the entanglement entropy satisfies an *area law for the entanglement entropy*. This boundary ∂A of the region A is defined as

$$\partial A := \{j \in A : \exists k \in B \text{ with } \text{dist}(j, k) = 1\}, \quad (9)$$



For a one-dimensional system, this boundary consists of two sites only, in a cubic lattice of dimension $\mathcal{D} > 1$ it contains $\mathcal{O}(L^{\mathcal{D}-1})$ many sites. That is, ground states of gapped models are less entangled than they actually could be, in fact *much less* entangled. This will be key to the understanding for what follows. For a comprehensive review on area laws, see Ref. [2].

1.3.3 Proven instances of area laws

Such an area law has been proven to be valid for a number of cases:

- For any gapped one-dimensional system with a unique ground state [12]. The proof again relies on Lieb-Robinson bounds mentioned before, albeit used in a much more sophisticated way than in order to show the clustering of correlations. This proof has in the meantime been significantly tightened using ideas of the *detectability lemma* [13].
- For gapped *free bosonic and fermionic models*, so for models where the Hamiltonian can be written as a quadratic polynomial of bosonic or fermionic annihilation and creation operators, the area law is known in fact for arbitrary lattices in any dimension [14, 15].
- For free bosonic models, the area law is even known to hold for critical, gapless models for cubic lattices of dimension $\mathcal{D} > 1$ [16].
- For some classes of states such as instances of so-called *graph states* [17] the area law holds by construction.
- Most importantly for the purposes of these lecture notes, *matrix product states* in one-dimensional systems and *projected entangled pair* states for higher-dimensional systems also satisfy an area law [18]. As we will see in a minute, this insight is not a detail at all: It is at the very heart of the insight why gapped quantum many-body systems can actually be numerically simulated using tensor network states.

- Once one has identified a system that satisfies an area law in higher-dimensions, one still obtains an area law for local Hamiltonians that are in the same quantum phase. This has been shown by making use of ideas of *quasi-adiabatic evolution*, Lieb-Robinson bounds and bounds to the generation of entanglement via local dynamics [19, 20].

Having said that, it should be clear that these findings can be seen as a large body of evidence that gapped many-body systems generically have the property that the ground state satisfies an area law.

1.3.4 Violation of area laws

For critical (gapless) one-dimensional models, the situation is quite distinctly different: The correlations decay too slowly in order to arrive at such a scaling of the entanglement entropy and the area law is violated. Still, the corrections to an area law are small. Conformal field theory suggests that [21, 22] $S(\rho_A) = (c/3) \log(l/a) + C$ where c is the conformal charge, a the lattice spacing, and $C > 0$ a non-universal constant, so in fact

$$S(\rho_A) = \mathcal{O}(\log(|A|)). \quad (10)$$

(again in Landau's asymptotic notation): It is logarithmically divergent in the size of the subsystem. For free fermionic models in 1D, and also for the XX model [23, 24] and the non-isotropic instance, the XY model [25], the precise scaling of the entanglement entropy has been rigorously proven using the *Fisher-Hartwig machinery* to compute *Toeplitz determinants*, again finding a logarithmic divergence, confirming a scaling that has been numerically determined in Ref. [26]. For a review, again see Ref. [14]; for an excellent non-technical introduction specifically in entanglement entropies in quantum field theories, see Ref. [27].

How do entanglement area laws in higher-dimensional *critical models* scale? This question is still largely open. It is known that critical free-fermionic systems violate an area law: For a cubic lattice in \mathcal{D} dimensions, one has

$$S(\rho_A) = \mathcal{O}(L^{\mathcal{D}-1} \log L), \quad (11)$$

which is (slightly) more than an area law would suggest [16, 28, 29]. Critical bosons, in contrast, can well satisfy an area law, even if critical [16].

1.3.5 Other concepts quantifying entanglement and correlations

The entanglement entropy is a unique measure of entanglement for pure states according to some axiomatic characterisation, but this does not mean that there are not a number of other qualifiers that meaningfully capture the entanglement content. Importantly in the context discussed here, one may replace the von Neumann entropy $S(\rho_A)$ by other *Rényi-entropies*

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \log_2(\text{tr}(\rho_A^\alpha)), \quad (12)$$

for $\alpha \geq 0$. For $\alpha \searrow 1$, this quantity reduces to the von-Neumann entropy of ρ_A . One also has $S_\infty(\rho_A) = -\log_2 \|\rho_A\|$, where $\|\rho_A\|$ is the operator norm of ρ_A and $S_0(\rho_A) = \log_2 \text{rank}(\rho_A)$. These slightly more general entanglement entropies will become important for our understanding of grasping ground states of one-dimensional models in terms of entanglement entropies. Again, for any of the above proven instances, it is also true that a more general Renyi-entropic area laws holds true as well.

For mixed states, such as for degenerate ground states or thermal states, the entropy of entanglement is no longer physically meaningful to quantify the degree of entanglement. There are essentially two strategies one can still pursue: On the one hand, one can look at measures capturing all correlations, not only quantum correlations or entanglement. The most accepted measure quantifying correlations is the *quantum mutual information*, defined as

$$I(A, B) := S(\rho_A) + S(\rho_B) - S(\rho). \quad (13)$$

For every Gibbs state $e^{-\beta H} / \text{tr}(e^{-\beta H})$ of a local Hamiltonian for some inverse temperature $\beta > 0$ of a local Hamiltonian H , this quantity is known to again satisfy an area law [30], albeit with a prefactor that grows linearly in the inverse temperature. These bounds can also be exponentially tightened for specific models such as the XX model [31]. On the other hand, mixed state entanglement measures can be employed that still capture genuinely quantum correlations even in mixed-state situations when quantum and classical correlations are intertwined. One of the most prominent such entanglement measure is the so-called *negativity* [32–34], defined as

$$E(\rho) = \|\rho^{T_A}\|_1 - 1, \quad (14)$$

where $\|\cdot\|_1$ is the trace norm ($\|O\|_1 = \text{tr}(|O|)$ for an operator O) and ρ^{T_A} is the partial transpose of ρ , so the operator that is obtained by taking a partial transpose with respect to the tensor factor belonging to subsystem A . *Entanglement negativities* have been studied in several contexts [35–37]. Since the logarithmic version, called *logarithmic negativity* $\log_2 \|\rho^{T_A}\|_1$ is an upper bound to the entanglement entropy, such quantities have also extensively been used in the past to derive area laws for entanglement entropies, even for non-degenerate ground states.

1.3.6 Entanglement spectra

The entropy of entanglement is just a single number, but it should be rather obvious that more detailed information is revealed when the entire spectrum of ρ_A is considered. In fact, the collection of all Renyi entropies of ρ_A gives precisely the same information as the spectrum of ρ_A itself. Given a state ρ_A , it is meaningful to consider the *entanglement Hamiltonian* H_A for which $\rho_A = e^{-H_A}$. In fact, the full entanglement spectrum (so the spectrum of ρ_A) reveals a lot of important information about the quantum many-body system and is receiving a significant attention in the context of *topological systems* and *boundary theories* [38–41].

1.4 The notion of the ‘physical corner of Hilbert space’

1.4.1 The exponentially large Hilbert space

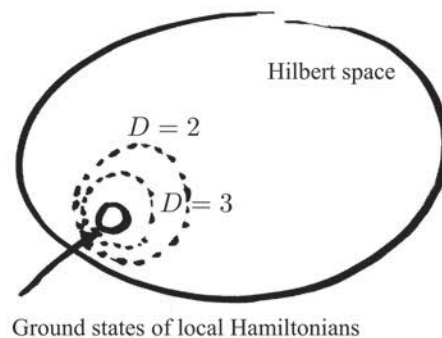
We have seen that ground states of gapped quantum many-body models exhibit little entanglement, in fact much less than they could feature. Interactions are short ranged, which not only means that correlations decay strongly, but also that there is very little entanglement present in the above sense. This is a basic, but at the same time very important, observation: It appears in this light that natural ground states (and Gibbs states) seem to explore only a very tiny fraction of Hilbert respectively state space that would in principle be available. Let us not forget that Hilbert space is exceedingly big: For a spin system of size n and local dimension d , the dimension scales as

$$\dim(\mathcal{H}) = \mathcal{O}(d^n). \quad (15)$$

It should be clear that already for moderately sized systems, state vectors can no longer be stored on a computer in order to numerically solve the system in exact diagonalisations (naively requiring $\mathcal{O}(d^{3n})$ operations). Surely, one can and must heavily exploit symmetries and sparsity patterns in the Hamiltonian to reduce the effective subspace that has to be considered, and then for moderately sized systems, exact diagonalisations can provide impressive results [42]. In any case, needless to say, one will encounter a scaling of the dimension of the relevant subspace that is exponential in the system size.

1.4.2 Small subset occupied by natural states of quantum many-body models

The key insight here is that the pure state exhibiting low entanglement in the sense of satisfying an area law constitute a very small subset of all pure states. What is more, this subset can be well approximated by tensor network states. In the end, the reason for tensor network methods to provide such powerful tools is rooted in the fact that natural ground states satisfy area laws (or show small violations thereof). In this sense, one might well say that the exponentially large Hilbert space ‘is a fiction’, and the ‘natural corner of Hilbert space’ constitutes only an exceedingly small fraction thereof. This somewhat wordy notion will be made more precise in a minute. One should not forget, after all, that not only ground states occupy a tiny fraction of Hilbert space, but the same holds true for all efficiently preparable quantum states: Not even a quantum computer could efficiently prepare all states [43], in fact a far cry from that:



2 Matrix product states

We now turn to exploiting this insight when grasping quantum many-body states in terms of tensor network states. We start by introducing a commonly used and quite illustrative graphical notation. We will then discuss in great detail the concept of a matrix product state which features in the highly successful density-matrix renormalisation group (DMRG) algorithm [1, 44–47]. It is worth noting that the history of this concept is actually quite remarkable. It did appear several times independently in the literature: *Finitely correlated states* as formulated in an algebraic picture in the language of mathematical physics [48] can be viewed as translationally invariant infinite matrix product states. In the same year as finitely correlated states were proposed, they independently emerged implicitly in the seminal work on DMRG by Steve White [44] in the context of condensed-matter physics – even if it took until much later until the close connection was spotted [46, 49]. In the meantime, the DMRG method is routinely explained in terms of matrix product states [1], a mindset that we will also follow. Yet independently, the concept of a *tensor train decomposition* [50] emerged in the mathematics literature, which was again found to be essentially equivalent to the concept of a matrix product state.

2.1 Preliminaries

2.1.1 Graphical tensor notation

A *tensor* can be represented as a multi-dimensional array of complex numbers. The dimensionality of the array required to represent it is called the *order* of the tensor. A scalar can be viewed as a tensor of order 0, a vector is seen as a tensor of order 1, a matrix would be tensor of order 2, and so on. We will make extensive use of the graphical notation that goes back to Penrose to represent tensors: We will graphically represent tensors as boxes, having a number of edges defined by the order of the tensor. This is then how a scalar looks like,



these are vectors and dual vectors,



and this



corresponds to a matrix. A contraction of an index amounts to summing over all possible values an index takes corresponding to a shared edge. For example, a matrix product $A = BC$ of matrices $A, B, C \in \mathbb{C}^{N \times N}$ amounts to

$$C_{\alpha,\beta} = \sum_{\gamma=1}^N A_{\alpha,\gamma} B_{\gamma,\beta}, \quad (16)$$

so here the common index γ is being contracted. Again, we can graphically represent this as

$$\boxed{A} \boxed{B} = \boxed{C}$$

The trace is a contraction of two indices of the same tensor, graphically



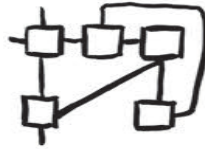
and a partial trace is



A scalar product looks like this:



The beauty and ease of this picture should be rather obvious.² An index that is not contracted is naturally called an *open index*. A *contraction of a tensor network*, like this one

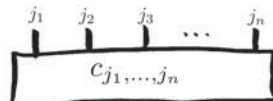


amounts to contracting all indices that are not open. Where would one best start with such a contraction? Indeed, the order very much matters as far as the complexity of the problem is concerned. The scaling of the effort of contraction in the dimension of the involved tensors is highly dependent on the contraction order, and to find the optimal order of pairwise contractions is a computationally hard problem in its own right. In practice, one often finds good contraction orders by inspection, however.

We now turn to the graphical representation of what we are mainly interested in, namely state vectors of quantum many-body spin-systems with n degrees of freedom. An arbitrary state vector $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$

$$|\psi\rangle = \sum_{j_1, \dots, j_n=1}^d c_{j_1, \dots, j_n} |j_1, \dots, j_n\rangle = \sum_{j_1, \dots, j_n=1}^d c_{j_1, \dots, j_n} |j_1\rangle \otimes \dots \otimes |j_n\rangle \quad (17)$$

with coefficients $c_{j_1, \dots, j_n} \in \mathbb{C}$ for all indices can be represented by



so by a box with n edges (sometimes also called ‘physical edges’ for obvious reasons).

²We will take a pragmatic viewpoint here and will swipe some mathematical fine print concerning such graphical tensor network representations of tensors over complex vector spaces under the rug which we should not too much worry about, however.

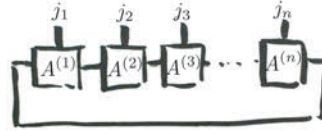
2.2 Definitions and preparations of matrix product states

2.2.1 Definition for periodic boundary conditions

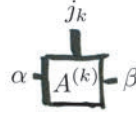
The definition of matrix product states takes the above tensor and ‘breaks it down’ to smaller components that are being contracted. A *matrix product state* [48, 51] of ‘bond dimension’ D (with periodic boundary conditions) is a pure state with a state vector of the form

$$c_{j_1, \dots, j_n} = \sum_{\alpha, \beta, \dots, \omega=1}^D A_{\alpha, \beta; j_1}^{(1)} A_{\beta, \gamma; j_2}^{(2)} \dots A_{\omega, \alpha; j_n}^{(n)} = \text{tr}(A_{j_1}^{(1)} A_{j_2}^{(2)} \dots A_{j_n}^{(n)}), \quad (18)$$

where the trace and the matrix product are taken over the contracted indices, leaving the physical indices j_1, \dots, j_n open. For a fixed collection of physical indices, the coefficients are hence obtained by considering matrix products of matrices, hence ‘matrix product state’. In a graphical notation, this can be represented as



That is to say, each individual tensor is represented as



and via contraction one arrives at the above expression. The line connecting the end tensors reflects the trace in the above expression. This graphical notation will remain very handy in what follows.

So what is D , the bond dimension? As such, it does not have a direct physical correspondence; this parameter can be viewed as a ‘refinement parameter’. It will also soon become clear why it is called a bond dimension and we will turn to its significance in a minute. Matrix product states constitute the, in many ways, most important instance of a tensor network state. They are of key importance both in analytical approaches as well as in numerical ones, most prominently in the density-matrix renormalisation group approach. Since we will frequently refer to such states, we will from now on commonly abbreviate them as MPS.

2.2.2 Variational parameters of a matrix product state

We note a first important property of a matrix product state: It is described by very few numbers. While a general state vector of a system composed of n spin- d systems is defined by $\mathcal{O}(d^n)$ real parameters, an MPS of bond dimension D can be represented by $\mathcal{O}(ndD^2)$ real parameters. For constant D , this is linear in n , as opposed to exponential in n : so this ansatz gives rise to a drastic reduction of the number of variational parameters, to say the least. At the same time it is true that D can be taken large enough that every state vector of a finite system can be

represented as an MPS, if one allows D to grow exponentially in n as well. Yet, this is actually not the main point of the definition of a matrix product state.

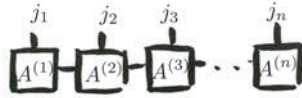
The key insight – one that should become increasingly clear – is that already for small bond dimension D , an MPS is an extraordinarily good approximation of natural states emerging in physical systems. The larger the bond dimension, so the ‘refinement parameter’ D , the larger is the set of states that can be represented, and hence usually the quality of the approximation of natural states. If one takes $D = 1$, then the above matrices merely become complex numbers and one obtains a product state, in a variational set that is sometimes referred to as a *Gutzwiller* variational state, a variant of a *mean-field approach*.

2.2.3 Matrix product states with open boundary conditions

The above expression corresponds to matrix product states for periodic boundary conditions. For open boundary conditions, the matrix $A^{(1)}$ is taken to be no longer a matrix from $\mathbb{C}^{D \times D}$, but $A^{(1)} \in \mathbb{C}^{1 \times D}$ so as a row vector. Similarly $A^{(n)} \in \mathbb{C}^{D \times 1}$ so it is a column vector. Then the expression becomes

$$c_{j_1, \dots, j_n} = \sum_{\alpha, \dots, \omega=1}^D A_{\alpha; j_1}^{(1)} A_{\beta, \gamma; i_2}^{(2)} \dots A_{\omega; j_n}^{(n)} = A_{j_1}^{(1)} A_{j_2}^{(2)} \dots A_{j_n}^{(n)}, \quad (19)$$

and the graphical expression



2.2.4 Area laws and approximation with matrix product states

What is the significance of area laws in the context of matrix product states? It is easy to see that for any subset A of consecutive sites of the lattice $S(\rho_A) = \mathcal{O}(\log(D))$ for a matrix product state, so the entanglement entropy is bounded from above by a constant in n . That is to say, MPS satisfy an area law. The behaviour of the entanglement scaling is therefore the same for matrix product states as for ground states of gapped models. But indeed, an even stronger converse statement is true: Every state that satisfies an area law can be efficiently approximated by a matrix product state.

There is a bit of fine-print associated with this statement: On the one hand, the precise wording of this statement makes use of Renyi entropic entanglement entropies as discussed above, and not the more standard entanglement entropies based on the von-Neumann entropy. (Yet, for all situations where a detailed understanding has been reached, this does not make a difference anyway, so we can be rather relaxed about it.) On the other hand, any such statement is surely one about a family of states of lattice systems of increasing system size, rather than for a single state. So precisely, if for a family of state vectors $|\psi_n\rangle$ there exist constants $c, C > 0$ such that for all $0 < \alpha < 1$ the Renyi entropies of the reduced state of any subsystem A of the one-dimensional system satisfy

$$S_\alpha(\rho_A) \leq c \log(n) + C, \quad (20)$$

then it can be efficiently approximated by an MPS (so the bond dimension will have to grow polynomially with n , the system size, and $1/\varepsilon$, where $\varepsilon > 0$ is the approximation error).

2.2.5 Preparation from maximally entangled pairs

There are two alternative convenient descriptions of MPS. The first one originates from a hypothetical preparation from maximally entangled states. This prescription explains why MPS can also be captured as ‘projected entangled pair states’ for one-dimensional spin chains. Let us assume, for that purpose, that each site of a one-dimensional spin chain is not composed of a single d -dimensional system, but in fact of two ‘virtual systems’ of dimension D . Each one of the pair of virtual systems is *maximally entangled* with the respective neighbouring system, which means that the state of this pair is described by the state vector

$$|\omega\rangle = \sum_{j=1}^d |j, j\rangle. \quad (21)$$

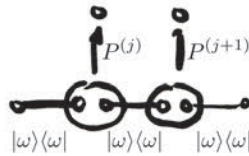
Such a pure state is indeed ‘maximally entangled’ in that the entanglement entropy of each subsystem takes the maximum value $\log_2(D)$. These entangled states are referred to as ‘bonds’, further motivating the above term of a ‘bond dimension’. On this system, we apply linear maps $P^{(j)} : \mathbb{C}^D \otimes \mathbb{C}^D \rightarrow \mathbb{C}^d$ for each of the sites $j = 1, \dots, n$, projecting two D -dimensional systems to a single system of the physical dimension d . We can hence capture the state vector as

$$|\psi\rangle = (P^{(1)} \otimes \dots \otimes P^{(n)}) |\omega\rangle^{\otimes(n-1)}. \quad (22)$$

This prescription, even though it may look somewhat baroque at this point, surely defines a valid state vector in $(\mathbb{C}^d)^{\otimes n}$. The claim now is that this is an alternative description of a matrix product state. How can that be? Let us write each of the the linear maps as

$$P^{(j)} = \sum_{k=1}^d \sum_{\alpha, \beta=1}^D A_{\alpha, \beta; k}^{(j)} |k\rangle \langle \alpha, \beta| \quad (23)$$

for $j = 1, \dots, n$, for periodic boundary conditions, graphically



Then a moments thought reveals that we again end up in a state vector with coefficients precisely of the form of Eq. (18). So the matrices that define the linear projection reappear in a different role in the definition of the matrix product state. An interesting side remark is that in this picture it is also particularly clear that MPS satisfy area laws, with $\log_2(D)$ being the maximum value the entanglement entropy can take. This picture will be particularly intuitive when generalising the idea of MPS to higher dimensional physical systems.

2.2.6 Sequential preparation picture

The second alternative convenient description of MPS relates to a sequential preparation of quantum states, an idea that was implicitly already present in Ref. [48]. Here, one starts off in a spin chain of local dimension d prepared in $|0\rangle^{\otimes n}$ and lets a quantum system of dimension D sequentially interact with each of the n constituents. At the end, one makes sure that the system is disentangled. It turns out that the state vectors generated in this way are exactly the MPS with open boundary conditions (see, e.g., Ref. [51]). More specifically, let $\sum_{j=1}^d \sum_{\alpha,\beta=1}^D A_{\alpha,\beta;j}^{(k)} |\alpha, j\rangle \langle \beta, 0|$ be an operation on $\mathbb{C}^D \otimes \mathbb{C}^d$ with

$$\sum_{j=1}^d (A^{(k)})_j^\dagger A_j^{(k)} = \mathbb{I} \quad (24)$$

(we will see below that this can always be chosen to be true) for each k , then one obtains an MPS with open boundary conditions of the form as in Eq. (19). This construction is interesting in many ways: To start with, this procedure gives rise to an efficient preparation of MPS, and there are several physical systems where one can readily think of systems sequentially interacting in this way (for example for atoms passing through cavities). In fact, in a slightly different language, MPS are discussed in the context of *quantum memory channels*, where the memory is encoded in the D -dimensional system passed from one site to the other. The second insight is that this sequential interaction picture plausibly motivates the exponential decay of correlation functions that we will learn about soon: All quantum correlations have to be mediated to the D -dimensional ‘ancilla’ system that is sequentially interacting.

2.2.7 Translationally invariant matrix product states

In the above definition, we have taken all matrices to be different. Of course, we can also in a translationally invariant ansatz choose them to be all the same, so take for periodic boundary conditions

$$A_{\alpha,\beta;k}^{(j)} = A_{\alpha,\beta;k} \quad (25)$$

for all $\alpha, \beta = 1, \dots, D$, all $k = 1, \dots, d$ and all sites $j = 1, \dots, n$. Such translationally invariant MPS make a lot of sense in analytical considerations, and obviously capture translationally invariant models well. They are specifically important when considering the thermodynamic limit $n \rightarrow \infty$. In numerical considerations, it is often advisable to break the symmetry and use different matrices per site even if the Hamiltonian as such is translationally invariant.

2.2.8 Successive Schmidt decompositions

The canonical form of an MPS can also be reached by making use of a *successive Schmidt decomposition*. This was first highlighted in Ref. [52]. We will be brief here, but explain the basic idea: Generally, a Schmidt decomposition of a state vector $|\psi\rangle \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ of a system

consisting of two parts A and B can be written as

$$|\psi\rangle = \sum_{j=1}^{\min(d_A, d_B)} \lambda_j |\psi_j^{(A)}\rangle \otimes |\psi_j^{(B)}\rangle, \quad (26)$$

with suitable orthonormal bases $\{|\psi_j^{(A)}\rangle\}$ and $\{|\psi_j^{(B)}\rangle\}$ of the respective Hilbert spaces, called left and right Schmidt vectors and $\lambda_j \geq 0$ for all j . Why is this remarkable? Because there is only a single sum, not a double sum. One can indeed now arrive at the canonical form in a one-dimensional MPS by starting from the left side and performing the Schmidt decomposition between site $\{1\}$ and the complement $\{2, \dots, n\}$ in V . Then one can expand the left Schmidt vectors in the original basis and continue by performing a Schmidt decomposition of the right Schmidt vectors between $\{2\}$ and $\{3, \dots, n\}$ and so on, to arrive at the normal form.

2.3 Computation of expectation values and numerical techniques

2.3.1 Computation of expectation values

How can we compute expectation values of MPS? Of course, the entire idea of a tensor network state only makes sense if we have a handle on meaningfully (at least approximately) computing expectation values $\langle \psi | O | \psi \rangle$ of local observables O . At this point, we have good reasons to hesitate, however, and to become a bit worried. The fact that we can describe an MPS by a few parameters alone does *not* necessarily imply that we can also efficiently compute the expectation value. For example, there are operations known, such as the permanent of a matrix, that cannot be computed in time polynomial in the dimension of the matrix (permanent is in the complexity class $\#P$).

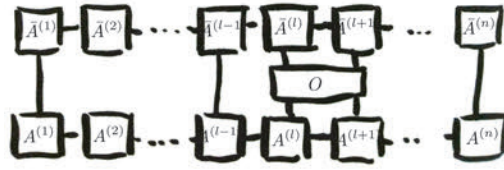
But let us see how far we get: Let us assume that O is a local term that is supported on neighbouring sites $l, l+1$ only, so

$$O = \sum_{j_l, j_{l+1}=1}^d \sum_{k_l, k_{l+1}=1}^d O_{j_l, j_{l+1}; k_l, k_{l+1}} |j_l, j_{l+1}\rangle \langle k_l, k_{l+1}|. \quad (27)$$

We suppress an index specifying the support. It should be clear that the same strategy can be applied to local terms with larger locality regions, so let us stick to nearest neighbour interaction terms for simplicity of notation. We pick open boundary conditions (but not necessarily a translationally invariant ansatz) to render the discussion more transparent. We now start from

$$\langle \psi | O | \psi \rangle = \sum_{j_1, \dots, j_n=1}^d \sum_{k_1, \dots, k_n=1}^d \bar{c}_{k_1, \dots, k_n} c_{j_1, \dots, j_n} \delta_{j_1, k_1} \cdots \delta_{j_{l-1}, k_{l-1}} O_{j_l, j_{l+1}; k_l, k_{l+1}} \delta_{j_{l+2}, k_{l+2}} \cdots \delta_{j_n, k_n}. \quad (28)$$

This expression looks quite messy. Resorting to the graphical notation, it can be more transparently be represented as

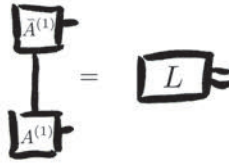


Naively formulated, we would clearly encounter $2n$ independent sums. In the light of this observation, one would hence be tempted to think that an effort exponential in n is needed to compute expectation values. Needless to say that much of the advantage of a tensor network state would disappear.

It is one of the key insights that expectation values of local observables can nevertheless be efficiently computed – one only needs to contract the tensor network in a smart order. Let us remind ourselves that the contraction effort is not independent on the actual order by which the contraction is performed. We start by investigating the left hand side of the tensor network: We can contract one index and write for the *left boundary condition*

$$L_{\alpha,\beta} := \sum_{j=1}^d A_{\alpha;j}^{(1)} \bar{A}_{\beta;j}^{(1)}, \quad (29)$$

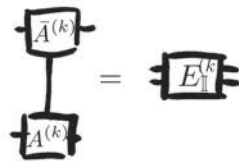
graphically represented as



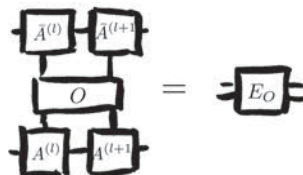
Now this again gives rise to a vector from $\mathbb{C}^{1 \times D^2}$: We have ‘doubled’ the indices and hence encounter an edge associated with dimension D^2 instead of D . We can now proceed as before, again contracting physical indices. In this way, we arrive at the *transfer operator* $E_{\mathbb{I}}^{(k)}$. This operator has components

$$(E_{\mathbb{I}}^{(k)})_{\alpha,\beta;\gamma,\delta} := \sum_{j=1}^d A_{\alpha,\beta;j}^{(k)} \bar{A}_{\gamma,\delta;j}^{(k)}. \quad (30)$$

At this point, the graphical notation seems straightforward



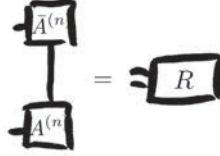
We can progress until we come to the sites on which the local observable O is supported. But of course, we can still contract all physical indices and treat it as one new tensor, to get E_O



Finally, the *right boundary condition* can be captured as

$$R_{\alpha,\beta} := \sum_{j=1^d} A_{\alpha;j}^{(n)} \bar{A}_{\beta;j}^{(n)}, \quad (31)$$

graphically



So the above graphical representation can also be sequentially read as a representation of the expression

$$\langle \psi | O | \psi \rangle = L E_{\mathbb{I}}^{(2)} E_{\mathbb{I}}^{(3)} \dots E^{(l-1)} E_O E_{\mathbb{I}}^{(l+2)} \dots E_{\mathbb{I}}^{(n-1)} R. \quad (32)$$

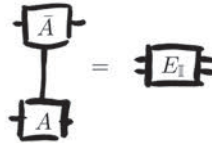
Of course, not only all these operators can be efficiently computed, but also the product be performed with the same effort as it is needed to multiply a vector from \mathbb{C}^{D^2} with matrices from $\mathbb{C}^{D^2 \times D^2}$, namely $\mathcal{O}(dD^4)$. Since there are n sites, the total effort to compute a single expectation value can be bounded by $\mathcal{O}(ndD^4)$. For a local Hamiltonian with $\mathcal{O}(n)$ terms H this amounts to an effort $\mathcal{O}(n^2D^4)$ to compute $\langle \psi | H | \psi \rangle$. That is to say, one can efficiently compute expectation values of local Hamiltonians. It is easy to see that this can be further improved to $\mathcal{O}(nD^3)$, by using an appropriate gauge, to be specified below, and altering the contraction order still.

2.3.2 Decay of correlations in infinite matrix product states

We have seen above that MPS capture ground states of local one-dimensional gapped models well. As such, one should expect that also common features of such models are appropriately reflected. In particular, one should expect correlation functions to decay exponentially with the distance in the lattice. In fact, the object we need in order to see this is the transfer operator encountered earlier. We stick to the situation of a infinite translationally invariant MPS, so the transfer operator

$$E_{\mathbb{I}} = \sum_{j=1}^d (A_j \otimes \bar{A}_j), \quad (33)$$

graphically



will not carry a site index. We will consider correlation functions between two sites that we label A and B as above. The observables supported on A and B are again O_A and O_B . In line with the above definitions, we set

$$E_{O_A} = \sum_{j,k=1}^d \langle k | O_A | j \rangle (A_j \otimes \bar{A}_k), \quad (34)$$

graphically

$$\begin{array}{c} \boxed{\bar{A}} \\ \boxed{O_A} \\ \boxed{A} \end{array} = \boxed{E_{O_A}}$$

and similarly for O_B . Then the correlation function is obtained as

$$\langle O_A O_B \rangle = \frac{\text{tr}(E_{O_A} E_{\mathbb{I}}^{\text{dist}(A,B)-1} E_{O_B} E_{\mathbb{I}}^{n-\text{dist}(A,B)-1})}{\text{tr}(E_{\mathbb{I}}^n)}. \quad (35)$$

We are now specifically interested in the limit $n \rightarrow \infty$ in *infinite translationally invariant MPS*. We have that

$$E_{\mathbb{I}}^k = |r_1\rangle\langle l_1| + \sum_{j=2}^{D^2} \lambda_j^k |r_j\rangle\langle l_j|, \quad (36)$$

where $\lambda_1 = 1, \lambda_2, \dots, \lambda_{D^2}$ are the eigenvalues of $E_{\mathbb{I}}$ in non-increasing order and $|r_j\rangle$ and $\langle l_j|$ the respective right and left eigenvectors. To assume that the largest eigenvalue $\lambda_1 = \|E_{\mathbb{I}}\| = 1$ does not restrict generality – this merely amounts to a rescaling. We also assume that this is the unique eigenvalue that takes the value 1. Then, in the limit $n \rightarrow \infty$,

$$\langle O_A O_B \rangle = \langle l_1 | E_{O_A} E_{\mathbb{I}}^{\text{dist}(A,B)-1} E_{O_B} | r_1 \rangle. \quad (37)$$

This is nothing but

$$\langle O_A O_B \rangle = \langle l_1 | E_{O_A} | r_1 \rangle \langle l_1 | E_{O_B} | l_1 \rangle + \sum_{j=2}^{D^2} \lambda_j^{\text{dist}(A,B)-1} \langle l_1 | E_{O_A} | r_j \rangle \langle l_j | E_{O_B} | l_1 \rangle, \quad (38)$$

where the first term can be identified as $\langle O_A \rangle \langle O_B \rangle$. This means that $|\langle O_A O_B \rangle - \langle O_A \rangle \langle O_B \rangle|$ decays exponentially in the distance $\text{dist}(A, B)$, and the correlation length $\xi > 0$ is given by the ratio of the second largest λ_2 to the largest λ_1 (here taken to be unity) eigenvalue of $E_{\mathbb{I}}$, so

$$\xi^{-1} = -\log |\lambda_2|. \quad (39)$$

This is a very interesting observation: The decay of correlations is merely governed by the spectral gap between the two largest eigenvalues of the transfer operator. All other details of the transfer operator do not matter asymptotically as far as the decay of correlations is concerned. This also means that whenever this gap is not vanishing, correlation functions always decay exponentially. Positively put, this may be taken as yet another indication that MPS represent ground states of gapped models well (for which correlation functions are similarly decaying). Higher order correlation functions of the form $\langle O_A O_B O_C \rangle$ and so on can also be efficiently computed from the MPS representation in the same way. There is an interesting structural insight related to this: In order to fully specify an MPS of any bond dimension, generically, the collection of all correlation functions of order up to merely three need to be specified [53], and not of all orders, as one would naturally expect.

In fact, MPS cannot represent algebraically decaying correlation functions, even though one should not forget that for sufficiently large bond dimension, MPS can well approximate states with algebraically decaying correlation functions well. One might be tempted to think that this is different whenever the gap is vanishing, meaning that whenever $\lambda_2 = 1$ and $\lambda_1 = 1$ are degenerate. This is not so, however. Then one rather encounters constant contributions to correlation functions (we will learn about an example of this form in form of the GHZ state below).

2.3.3 Computation of scalar products between matrix product states

It is an interesting exercise to verify that the scalar product of two different (non-translationally invariant, open boundary condition) MPS of bond dimension D can be computed with an effort $\mathcal{O}(ndD^3)$. This fact can be quite helpful in practical computations.

2.3.4 Density-matrix renormalisation method in several invocations

The workhorse of the numerical study of strongly correlated one-dimensional systems is the *DMRG method*, introduced in the seminal Ref. [44]. Albeit this was not the way this method has originally been formulated, it has become clear [46,49] that it can be viewed as a variational method over MPS: In one way or the other, one varies over MPS state vectors $|\psi\rangle$ of a given bond dimension until for a given local Hamiltonian $H = \sum_{j \in V} h_j$ a good approximation of

$$\min \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \quad (40)$$

is reached. We can describe the MPS by polynomially many (in fact in n linearly many) parameters that can be stored in a computer, and we can efficiently compute expectation values. Since the optimisation over all parameters at once amounts to a non-convex global optimisation problem (and is hence infeasible), this task is broken down to a sequential updating of the matrices $\{A_{j_k}^{(k)}\}$ of the MPS. For example, starting from randomly picked matrices in the MPS, if one holds all matrices except those $\{A_1^{(j)}, \dots, A_d^{(j)}\}$ of a site j fixed, then one can write the optimisation problem of Eq. (40) as a minimisation over

$$E := \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle A^{(j)} | K_1 | A^{(j)} \rangle}{\langle A^{(j)} | K_2 | A^{(j)} \rangle} \quad (41)$$

with $|A^{(j)}\rangle$ denoting the vectorized forms of the matrices and K_1 and K_2 being the kernels of the respective quadratic forms. This is not only a convex quadratic optimisation problem, but in fact an eigenvalue problem $K_1 |A^{(j)}\rangle = E K_2 |A^{(j)}\rangle$. In this way, by ‘sweeping through the lattice’, going from one site to the next and coming back, until convergence is reached, one can find ground state properties essentially up to machine precision. In practice, often surprisingly few sweeps are needed to reach a stable minimum, even if one starts off with random MPS. See Refs. [1, 3, 45] for reviews.³

³As a side remark, strictly speaking, it is not guaranteed by this procedure that one really obtains the global minimum when performing local variations. In fact, practically one may get stuck, and it can be beneficial to insert

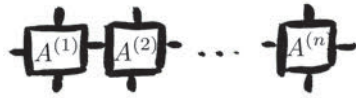
Having said that, there are, needless to say, many different methods of how to proceed and many different ways in which one can improve this basic method. To start with, a clever use of the gauge (as we will discuss later) is crucial to arrive at a practical implementation avoiding *ill-conditioned matrices* along the way. Then, one does not have to vary over one set of matrices per step, but can vary over pairs of matrices, in a *double variational site*, leading in particular to a faster convergence and a better error control. One can employ instances of *time-evolving block decimation (TEBD)* [57–59] in imaginary-time evolution (see the next subsection), or the *time-dependent variational principle* [60] avoiding Trotter errors. Some variants of DMRG avoid errors from finite-size scaling by directly referring to infinite MPS (possibly with a broken symmetry with a finite period), such as the *iDMRG method* [61, 62] or *iTEBD* [63].

2.3.5 Matrix-product operators and mixed states

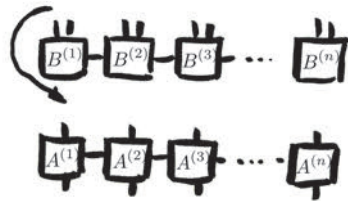
Tensor network states can not only represent pure states, but mixed quantum states as well. A *matrix product operator* $O \in \mathcal{B}((\mathbb{C}^d)^{\otimes n})$ relates to a tensor network

$$O = \sum_{j_1, \dots, j_n=1}^d \sum_{k_1, \dots, k_n=1}^d \text{tr}(A_{j_1, k_1}^{(1)} \dots A_{j_n, k_n}^{(n)}) |j_1, \dots, j_n\rangle \langle k_1, \dots, k_n|, \quad (42)$$

These operators contain mixed quantum states (and also other operators which are not positive in the sense that eigenvalues can be negative; in fact, checking positivity of such an matrix product operator is not straightforward). Graphically they can be represented as



One can now devise algorithms that represent mixed states such as Gibbs states $e^{-\beta H} / \text{tr}(e^{-\beta H})$ in variants of the DMRG algorithm [64–66]. There is a second picture capturing mixed states, which are obtained from MPS with a special structure by taking a partial trace over the purifying degrees of freedom, here depicted in the form of an isomorphism between purified MPS and positive instances of MPO [67]



2.3.6 Time evolution

Time evolution algorithms provide numerical computations of expectation values

$$\langle O_A \rangle(t) := \langle e^{itH} O_A e^{-itH} \rangle \quad (43)$$

manually artificial fluctuations [44]. In practice, this is usually not much of a problem, however. The *computational complexity of actually finding the optimal MPS*, given a fixed family of Hamiltonians and a given bond dimension, has been addressed in Refs. [54–56].

for a local Hamiltonian H and a local observable O_A supported on several sites. Basically, there are two major strategies known to proceed here, as far as MPS simulations are concerned:

- On the one hand, one can decompose the Hamiltonian $H = H_1 + H_2$ into a sum of an even and an odd part, such that both Hamiltonians contain non-overlapping Hamiltonian terms only. One can then approximate e^{-itH} by $(e^{itH_1/k} e^{itH_2/k})^m$ for a suitable large $m \in \mathbb{N}$, with good control of errors. The *time-evolving block decimation (TEBD)* [57–59] and variants thereof make use of that idea.
- On the other hand, one can make use of the *time-dependent variational principle* [60], which relies on the variational manifold of uniform MPS mentioned above and which avoids Trotter errors altogether.

Such methods have provided significant new insights into *non-equilibrium dynamics of strongly correlated quantum systems* and have given new impetus to the study old questions of equilibration or thermalisation. Specifically, quenched quantum lattice models have been extensively studied, resulting in much progress on questions of far-from-equilibrium dynamics (see, e.g., Refs. [68–71]).

For short times, starting from clustering initial conditions, these methods again provide very reliable information about the evolution of the strongly correlated many-body systems. The reason is again rooted in the above mentioned Lieb-Robinson bounds: One can show that under local Hamiltonian dynamics, for any fixed time, an area law for Renyi entanglement entropies holds true [72]. Hence, an efficient approximation of the true time evolved states with an MPS is possible. The prefactor in the area law grows with time, however, leading to the situation that the entanglement effectively increases linearly in time in worst case [73]. That is, for long times, one cannot capture time evolution of quantum many-body systems with MPS: One hits the ‘barrier of entanglement growth’. So-called *folding methods* that contract the emerging tensors in a different way soften this problem to some extent [74]. Still, to grasp long time dynamics is infeasible, and it is one of the interesting open problems to see to what extent this challenge can be overcome.

2.4 Parent Hamiltonians, gauge freedom, geometry, and symmetries

At this point, it should have become clear that MPS are primarily meant to approximate natural states, specifically ground states of gapped one-dimensional local Hamiltonian models. Yet, a natural question that arises at this point is whether there are meaningful Hamiltonians that have an MPS as their exact ground state. We will now take a look at this question. Starting from this observation, we will hint at the insight that indeed, MPS (and more generally tensor network states) are by no means only powerful tools in numerical simulations, but also give rise to a very versatile tool in analytical considerations. Since so many questions can be fully analytically assessed (in contrast to many actual Hamiltonian problems), an entire research field has emerged of ‘studying condensed matter physics in the MPS world’. For example, complete classifications of quantum phases have been given [75–77], new instances of *Lieb-Schultz-Mattis*

theorems proven [78], *fractional magnetisation* considered [79] *phase transitions of arbitrary order* identified [80], or a ‘*Wick’s theorem*’ for MPS be formulated [53]. It will be beyond the scope of this book chapter to give a comprehensive overview of this development.

2.4.1 The AKLT and the Majumdar Gosh models

Surely, any product state is an MPS with bond dimension $D = 1$, so every state that has a product state as an exact ground state will provide an example of that sort. This is not very exciting yet, however. The most famous model that has an MPS ground state with a bond dimension different from $D = 1$ is the *AKLT model*, named after Affleck, Kennedy, Lieb, and Tasaki [81]. Here, the bond dimension is $D = 2$ while the local physical dimension is $d = 3$: It is a spin-1 chain. There are many ways to progress to see that this model indeed has an MPS as its ground state. One particularly transparent way makes use of the above projected entangled pair picture – historically, this is actually where this construction originates from. The linear projections are all the same for all sites and taken to be

$$P = \Pi_{S=1}(\mathbb{I} \otimes iY) \quad (44)$$

where $\Pi_{S=1}$ is the projection onto the spin-1 subspace of two sites, and Y is the Pauli matrix. This might look a bit awkward at first, but it is a sanity check that it takes a state defined on two spin-1/2 systems (with Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$) to one spin-1 state (defined on \mathbb{C}^3). The Pauli matrix is just added here for the convention that we always start in the maximally entangled state as defined in Eq. (21). So what this linear map essentially does is that it takes bonds prepared in singlets and projects them into the $S = 1$ subspace. This surely gives rise to a valid MPS with state vector $|\psi\rangle$.

Why does this help to construct a Hamiltonian that has this MPS as the ground state? We can simply take as the local Hamiltonian term $h_j = \Pi_{S=2}$, so that surely $h_j|\psi\rangle = 0$ for all j . That is to say, the Hamiltonian terms are the projections onto the $S = 2$ subspace. For the same reason,

$$H|\psi\rangle = \sum_j h_j|\psi\rangle = 0 \quad (45)$$

(mildly disregarding some fine-print on the boundary conditions). But we have that $h_j \geq 0$, so that each term has non-negative eigenvalues, which means that $|\psi\rangle$ must be a ground state vector. There is some fine print involved here, as strictly speaking, we have only seen that it constitutes a valid ground state vector, not really whether it is a unique one. This can be shown to be true, however, by identifying the Hamiltonian in Eq. (45) as the so-called parent Hamiltonian of the given MPS, a concept we will turn to later.

How does the resulting Hamiltonian then look like? A moment of thought reveals that it can be written as

$$h_j = \frac{1}{2}S^{(j)} \cdot S^{(j+1)} + \frac{1}{6}(S^{(j)} \cdot S^{(j+1)})^2 + \frac{1}{3}, \quad (46)$$

The matrices of the MPS are found to be $A_1 = X$, $A_2 = (X + iY)/\sqrt{2}$, $A_3 = -(X - iY)/\sqrt{2}$. In fact, one of the motivations to study the AKLT model is also the close resemblance to the

spin-1 Heisenberg model the Hamiltonian of which has local terms

$$h_j = JS^{(j)} \cdot S^{(j+1)} \quad (47)$$

for some $J \in \mathbb{R}$. This model is important for numerous reasons. It is also connected to the famous *Haldane conjecture*, which states that integer-spin anti-ferromagnetic Heisenberg chains are gapped [82].

Another important local Hamiltonian that has an exact MPS ground state is the one of the *Majumdar-Gosh model* [83], a nearest-neighbour spin-1/2 chain of local dimension $d = 2$ with Hamiltonian

$$H = \sum_j (2\sigma^{(j)} \cdot \sigma^{(j+1)} + \sigma^{(j)} \cdot \sigma^{(j+2)}), \quad (48)$$

where σ is the vector of Pauli matrices. It turns out that its ground state can be represented with matrices of bond dimension $D = 3$.

2.4.2 Gauge freedom and canonical forms

An MPS is uniquely defined by the matrices defining it, but the converse is not true: There is more than one set of matrices that give rise to the same pure state. Since

$$A_{j_k}^{(k)} A_{j_{k+1}}^{(k+1)} = A_{j_k}^{(k)} X X^{-1} A_{j_{k+1}}^{(k+1)} \quad (49)$$

for any $X \in Gl(D, \mathbb{C})$, graphically represented as

one can introduce between any pair of MPS matrices an invertible matrix. Choosing a specific X , or rather choosing such matrices for an entire MPS, amounts to choosing a so-called *gauge*. This insight, basic as it is, is very helpful in many circumstances. *Canonical forms*, so special forms of MPS that are particularly simple and which can be achieved by picking a suitable gauge, are very useful in analytical considerations. They are also helpful in numerical methods: For example, the above reduction of the effort when computing expectation values from the naive $\mathcal{O}(n^2 D^4)$ to $\mathcal{O}(n D^3)$ is also partially due to picking the right gauge. For MPS with open boundary conditions and bond dimension D , one can, for example, always pick a gauge in which

$$\sum_j A_j^{(k)} (A_j^{(k)})^\dagger = \mathbb{I}, \quad (50)$$

$$\sum_j (A_j^{(k)})^\dagger A_j^{(k-1)} A_j^{(k)} = A_j^{(k)}, \quad (51)$$

$$A^{(0)} = A^{(n)} = 1 \quad (52)$$

and each $A^{(k)} \in \mathbb{C}^{D \times D}$ for $k = 1, \dots, n-1$ is diagonal, positive, has full rank and unit trace. This can be shown by a successive singular value decomposition. For periodic boundary conditions, finding appropriate normal forms is more involved, see Ref. [51].

2.4.3 Injectivity and parent Hamiltonians

We have seen that the AKLT model has a unique MPS as its ground state, and so does the Majumdar-Gosh model. Indeed, these are examples of a more general framework that goes back to the seminal work of Ref. [48], where the question has been solved for infinite systems. Here, we ask, for a finite MPS, when is it the unique ground state of a gapped local Hamiltonian? The latter is usually called *parent Hamiltonian* of the MPS. Let us assume we have a MPS state vector in the canonical form of Subsection 2.4.2

$$|\psi\rangle = \sum_{j_1, \dots, j_n=1}^d A_{j_1}^{(1)} \dots A_{j_n}^{(n)} |j_1, \dots, j_n\rangle. \quad (53)$$

We assume now that we can group the constituents such that the grouped MPS, now no longer of n but of m new sites of larger local dimension, to get

$$|\psi\rangle = \sum_{j_1, \dots, j_m} B_{j_1}^{(1)} \dots B_{j_m}^{(m)} |j_1, \dots, j_m\rangle. \quad (54)$$

We require that each sets of matrices $\{B_{j_k}^{(k)}\}$ has the property that it generates the respective entire space of matrices.⁴ Then one can find local Hamiltonian terms $\{h_j\}$ each supported on $L + 1$ sites of the original lattice such that $|\psi\rangle$ is the unique ground state vector of $H = \sum_j h_j$. This is in fact a generalisation of the above idea that we encountered in the AKLT model (strictly speaking, there injectivity sets in at $k = 2$, so one would in principle arrive at a 3-local Hamiltonian, but one can show that this Hamiltonian and the given nearest-neighbour Hamiltonian of the AKLT model are in fact identical). The existence of a gap can for finite systems be shown in essentially the same way as for infinite finitely correlated states [48]. The idea that one can (under the above mild technical conditions) find a gapped local Hamiltonian of which the MPS is the unique ground state is a very powerful one in analytical uses of MPS.

2.4.4 Group symmetries

Significant progress has been made in recent years in the study tensor networks – and specifically MPS – under symmetry constraints. Here the emphasis is pushed from the study of symmetries in Hamiltonians to those of states, but there is a close connection (making use of the concept of a parent Hamiltonian of the previous subsection): If an MPS state is invariant under a representation of a group, then one can choose its parent Hamiltonian to be invariant under the same representation.

One of the key features of a translationally invariant state vector $|\psi\rangle$ on n sites is the symmetry group G under which it is invariant: This is the group for which

$$u_g^{\otimes n} |\psi\rangle = e^{i\phi_g} |\psi\rangle, \quad (55)$$

⁴Let $\{A_{j_k}^{(k)}, \dots, A_{j_{k+L-1}}^{(k+L-1)}\}$ correspond to the block called $B^{(s)}$ for some suitable s , then the above means that the map $\Gamma_L : Gl(D, \mathbb{C}) \rightarrow (\mathbb{C}^d)^{\otimes L}$ with $\Gamma_L : X \mapsto \sum_{j_k, \dots, j_{k+L-1}=1}^d \text{tr}(X A_{j_k} \dots A_{j_{k+L-1}}) |j_1, \dots, j_L\rangle$ is *injective* [51]. If for a state vector such an L can be found, then the state vector is called injective.

where $g \in G$ and u_g is a unitary representation on \mathcal{H} . It turns out that for translationally invariant MPS that fulfill the injectivity condition this symmetry is reflected also by a group symmetry in the tensors that define the MPS: The action of the unitary u_g on the physical index corresponds to an action of a V_g on the virtual level. More specifically, $u_g P = P(V_g \otimes \bar{V}_g)$ for the linear operators P defining the MPS ($P = P^{(j)}$ for all j in a translationally invariant ansatz). This picture has shed new light on the concept of *string order* [84]. It also plays an important role in the *classification of phases* [75–77], when two gapped systems are defined to be in the *same phase* if and only if they can be connected by a smooth path of gapped local Hamiltonians.

2.4.5 Manifold of matrix product states

There is an interesting geometrical structure associated with (translationally invariant) MPS. We have seen that there is a gauge freedom in MPS, leading to an over-parametrisation. Due to this redundancy in parametrisation, MPS have the structure of a principal fiber bundle. The bundle space corresponds to the entire parameter space, that is, the collection of all tensors associated with the physical sites. The base manifold, in turn, is embedded in the Hilbert space. This geometrical structure is fleshed out in detail in Ref. [85].

2.5 Tools in quantum information theory and quantum state tomography

2.5.1 Matrix product states in metrology

Many multi-particle entangled states that are interesting in the context of quantum information theory and metrology can also be represented as matrix product states. The well-known *Greenberger-Horne-Zeilinger (GHZ) state* with state vector

$$|\psi\rangle = (|0, \dots, 0\rangle + |1, \dots, 1\rangle)/\sqrt{2}, \quad (56)$$

for example, can be written as a MPS with bond dimension $D = 2$ and $A_1 = |0\rangle\langle 0|$ and $A_2 = |1\rangle\langle 1|$. For practical applications in metrology, GHZ states are too fragile with respect to noise, and other states which can also be represented as MPS are more useful [86].

2.5.2 Matrix product states in measurement based quantum computing

Cluster states are also MPS: These states are an important class of states in quantum information theory, most prominently featuring (in their two-dimensional instance) in ideas of *measurement-based quantum computing* [87, 88]: This is the idea of performing quantum computing without the need of actual unitary control over arbitrary pairs of constituents, but rather by sequentially (and adaptively) measuring single sites. Since read-out has to be done anyway at some stage even in the circuit model, this amounts to an appealing picture of quantum computing. All the entanglement ‘consumed’ in the course of the computation is already present in the initial, rather simple, resource state. Cluster states are an instance of the more general *graph states* [17], which constitute a helpful theoretical ‘laboratory’: They can be viewed as prepared in the

following way: One starts with preparing the vertices in a lattice V in $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and applying *controlled phase gates* $|j, k\rangle \mapsto |j, k\rangle e^{i\delta_{j,1}\delta_{k,1}\phi}$ to neighbouring systems; in the original definition, the phase $\phi = \pi$ is chosen.⁵ In one-dimension, a cluster state vector is (with obvious adaptations at the boundaries) the unique eigenvector of a set of mutually commuting *stabiliser operators*

$$K^{(j)} = Z^{(j-1)} X^{(j)} Z^{(j+1)} \quad (57)$$

for $j = 2, \dots, n-1$. It is left as an interesting and not very difficult exercise to the reader to find out how this state vector can be represented as an MPS with local dimension $d = 2$ and bond dimension $D = 2$.

In Ref. [88] *new models for measurement-based computing* have been proposed, taking the idea seriously that the matrices used in the parametrization of an MPS can be directly understood as quantum gates on a logical space. Indeed, this mindset gives rise to a wealth of novel models, an idea that has turned out to be fruitful since then. For example, resource states can be found exhibiting long-range correlations and variants of the ground state of the AKLT model can be taken to be resource states [90, 91]. In Ref. [92] a complete classification of *qubit wires* (spin systems allowing for a transport of quantum information) is given in an instance where a physically well-motivated class of universal resources can be fully understood, using ideas of classifications of quantum channels.

2.5.3 Localizable entanglement

The ideas of the previous subsection also relate to the concept of *localizable entanglement* [93]: This is the characteristic length scale with which two sites $A, B \in V$ can be entangled by making local projective measurements on all other sites $V \setminus \{A, B\}$. This length scale can be much longer than the traditional correlation length as in Eq. (6). In fact, there are gapped quantum Hamiltonians the unique ground state of which exhibits an infinite localisable entanglement, but a finite correlation length.

2.5.4 Matrix product states in quantum state tomography

The same reason why MPS (and again more general tensor network states) are so powerful in numerical approaches to problems in condensed matter physics render them also optimally suited for another purpose: For *quantum state (or process) tomography*. This is the important and natural task of estimating an unknown state (or process) from measurements. This is obviously one of the key tasks that experimentalists routinely face when performing experiments with precisely controlled quantum systems. It is beyond the scope of the present chapter to give a comprehensive overview over this important field of research. Still, from the above it

⁵In fact, even states that can be prepared by applying arbitrary non-local phase gates associated to any interaction graph applied to an arbitrary MPS can be efficiently contracted. This is possible by suitably defining transfer operators that absorb the phases in such a way that the long-range entanglement is not an obstacle to an efficient contraction. The schemes arising from this variational set of states are referred to as *renormalisation schemes with graph enhancement* [89]. Such states are efficiently contractable states strongly violating an area law.

should be clear where the insights developed here come in: In order to faithfully reconstruct an unknown pure generic many-body state of local dimension d and n sites from expectation values, one needs to know $\mathcal{O}(d^n)$ different expectation values. In order to reconstruct an unknown MPS, in sharp contrast, merely $\mathcal{O}(nD^2d)$ expectation values are needed, an exponential improvement. What is more, one can obtain the relevant information from learning suitable *reduced density operators* alone [94]. Similar ideas can also be applied to *quantum fields* and continuum systems, using the concept of continuous matrix product states that we will encounter later [53]. Without such tools, it seems that the certification of quantum experiments can soon no longer keep up with experimental progress with controlled quantum systems.

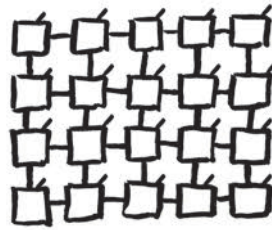
3 Higher-dimensional tensor network states

The idea of a tensor network is by no means confined to one-dimensional quantum systems. In fact, one would usually rather refer to an actual ‘tensor network’ if the topology is not that of a one-dimensional chain. We start with the higher-dimensional analogue of matrix product states and then turn to other approaches such as multi-scale entanglement renormalisation.

3.1 Higher-dimensional projected entangled pair states

3.1.1 Definition of projected entangled pair states

A *projected entangled pair state* (PEPS) [95], closely related to the older concept of a *tensor network state* [96–98], is the natural generalisation of an MPS to higher-dimensional systems. For a cubic lattice $V = L^{\mathcal{D}}$ for $\mathcal{D} = 2$ and open boundary conditions, the tensor network can be graphically represented as



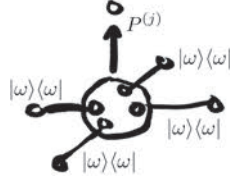
with the natural analogue for periodic boundary conditions on the torus. Similarly as before, each of the tensors – now five index tensors $A_{\alpha,\beta,\gamma,\delta;j}^{(k)}$ – can be chosen all different for all sites $k \in V$, with $\alpha, \beta, \gamma, \delta = 1, \dots, D$ and $j = 1, \dots, d$. But they can again also all be the same (possibly with the exception of the tensors at the boundary) in a translationally invariant ansatz class.

Again, one can formulate an actual projected entangled pair picture: Imagine again that each pair of physical sites in V shares with each nearest neighbour a maximally entangled state, as defined in Eq. (21), in the virtual degrees of freedom. To this state again a linear map $P^{(j)}$ is applied for each site, now (with variations at the boundary for open boundary conditions) a map

$P^{(k)} : (\mathbb{C}^D)^{\otimes 4} \rightarrow \mathbb{C}^d$, defined as

$$P^{(k)} = \sum_{\alpha, \beta, \gamma, \delta=1}^D \sum_{j=1}^d A_{\alpha, \beta, \gamma, \delta; j}^{(k)} |j\rangle \langle \alpha, \beta, \gamma, \delta|. \quad (58)$$

Again, one ‘projects entangled pairs’,



3.1.2 Properties of projected entangled pair states

A number of natural properties emerge that render this class of variational states a natural one:

- PEPS satisfy area laws, and it takes a moment of thought to see that the entanglement entropy of a subset A is bounded from above by $\mathcal{O}(L \log D)$ for $D = 2$: one simply needs to disentangle as many bonds as the boundary ∂A of A contains.
- Again, if the bond dimension D is large enough, then one can approximate (or for finite systems explicitly write out) every state as a PEPS.
- One can also again have exponential clustering of correlations. Interestingly, here already a difference emerges to MPS in one dimension: One can construct PEPS that have algebraically decaying correlations with $\text{dist}(A, B)$ between two sites or regions $A, B \subset V$ [18].

Such a strong statement on how well general states can be approximated with PEPS as it is available for MPS is lacking: One expects, however, that states satisfying area laws – so presumably ground states of gapped models – can be well approximated with PEPS with a small bond dimension D . Also, the body of numerical evidence available shows that this class of states indeed meaningfully describes strongly correlated quantum many-body systems.

3.1.3 Approximate contraction of projected entangled pair states

Again similarly to the situation before, one can define *transfer operators* $E_{\mathbb{I}}$, graphically represented as



and similarly E_{O_A} in case of the presence of a local observable. In contrast to MPS, PEPS cannot be efficiently contracted exactly, however. Strictly speaking, the contraction is contained in

the complexity class $\#P$, again giving rise to a computationally hard problem [99]. It is also intuitive why naive exact contractions cannot work: No matter what contraction order one picks, one will end up with an object having a number of edges that is linear in L . Therefore, computations such as the determination of expectation values of observables O_A are, differently from the situation in MPS, strictly speaking inefficient. This is no reason for panic, however: There are promising results on *approximate contraction techniques* that allow for a good practical description of gapped local two-dimensional Hamiltonians. For example, one can work oneself through the scheme and contract each row with the subsequent one: Before progressing, however, one approximates the bond dimensions of the new tensors by those the tensors had previously [95]. This is no longer exact, but feasible and efficient (the best known such scheme displays an effort of $\mathcal{O}(D^8)$ in the bond dimension). Another approximate way of contracting amounts to renormalising the cubic lattice to a new cubic lattice of smaller system size, and again approximates the obtained tensors of higher bond dimension by the previous ones [100, 101]. There has also been recent insight into why such approximation methods should be expected to provide good service, using ideas of *local topological order* [102].

3.1.4 Infinite methods

Again, the basic ideas have been generalised to the situation of having an infinite system to start with, to avoid the need for finite size scaling. A number of methods have been suggested, among them *boundary-MPS methods* [103], *corner transfer matrix* methods [96], as well as again *tensor coarse-graining methods* [100, 101]. These methods provide competitive numerical simulations of two-dimensional lattice models, for a review, see Ref. [104].⁶

3.1.5 Exact models

It goes without saying that again PEPS are not only a useful tool in numerical studies, but in analytical ones as well. *Cluster states* in two dimensions [87] are instances of PEPS, and so are a number of other classes of states important in quantum information theory. The models of Refs. [88, 90] for *measurement-based quantum computing* are also based on PEPS. The possibly most important Hamiltonian with PEPS as ground states is the *toric code Hamiltonian*

$$H = -J_a \sum_s A_s - J_b \sum_p B_p \quad (59)$$

defined on the edges of a two-dimensional cubic lattice, where $\{A_s\}$ and $\{B_p\}$ are *star* and *plaquette operators*, respectively, defined as

$$A_s = \prod_{j \in s} X^{(j)}, \quad B_p = \prod_{j \in p} Z^{(j)}, \quad (60)$$

⁶To be fair, one should add that at present, one-dimensional approaches based on matrix product states are still better developed than those based on higher-dimensional tensor networks. The ‘crime story’ of the precise nature of ground state of the spin-1/2 *Heisenberg anti-ferromagnet on the Kagome lattice* with local Hamiltonian terms as in Eq. (47) – it is a spin liquid ground state – has finally been resolved using DMRG and a ‘snake-like’ one-dimensional ordering of the tensors of the two-dimensional Kagome lattice, and not using an algorithm using PEPS or multi-scale renormalisation [105, 106].

so by the product of Pauli operators around a star or around a plaquette. On the one hand, this model is interesting in that it can be viewed as a lattice instance of a \mathbb{Z}_2 lattice gauge theory. On the other hand, it is the most prototypical quantum lattice model exhibiting *topological order*. It may hence by no surprise that the literature on this model is very rich, to say the least.

There are also models considered in the literature that can not be exactly solved by means of PEPS, but for which variational approaches of PEPS with few variational parameters already give good energies and significant physical insight into the model at hand, so have a somewhat analytical flavour. Such a mindset has been followed, e.g., in the study of *resonating valence bond wave-functions* [107].

3.2 Multi-scale entanglement renormalization

So far, we have been discussing tensor networks that had the same topology as the underlying physical lattice. Needless to say, there are good reasons to choose other topologies as well. A guideline is served by the criteria that (i) the tensor network should be described by polynomially many parameters, (ii) it should be efficiently contractible, either exactly or approximately, and (iii) the corresponding class of quantum states should be able to grasp the natural entanglement or correlation structure.

3.2.1 Tree tensor networks

Specifically, for critical models, one would expect a scale invariant ansatz class to be reasonable, one that reflects the scale invariant nature of the ground state. A first attempt in this direction is to think of tree tensor networks [108, 109]. For example, one can think of a *binary tree tensor network*: Here, one introduces a fictitious time in a renormalisation scheme where in each step, two sites are mapped into a single one by an *isometry*. At the top layer of this hierarchy, one prepares two systems in some initial pure state. This scheme has several of the above advantages: (i) It is described by polynomially many parameters, (ii) one can efficiently contract the network, and (iii) the states generated inherit the scale invariance of the ansatz class. There are also disadvantages: notably, there are sites in the physical lattice that are nearest neighbours; yet, in the tensor network they are only connected via the top layer in the hierarchy. Tree tensor networks with an entire symmetric subspace on the top layer and not only a single state naturally emerge as exact *ground states of frustration-free spin Hamiltonians* [110].

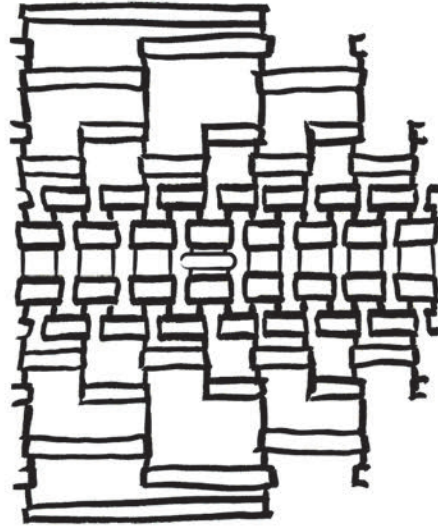
3.2.2 Multi-scale entanglement renormalisation

A generalisation has been suggested in Ref. [111], referred to as *multi-scale entanglement renormalisation (MERA)*. The idea can be viewed as a refinement of the binary tree mentioned above, to what is called a binary MERA. For clarity of notation, we consider one-dimensional chains. Again, one thinks of a tensor network having different *layers* or temporal steps. Let us imagine that $n = 2^T$, meaning that we think of T temporal layers, labeled $t = 1, \dots, T$. $t = 1$ corresponds to the physical layer, $t = T$ to the top layer. However, a new ingredient is

added. Each temporal step now consists of two elementary steps. One elementary step is again a renormalisation step, invoked by an isometry $I : \mathbb{C}^{d_j \otimes d_j} \rightarrow \mathbb{C}^{d_{j+1}}$ satisfying $I^\dagger I = \mathbb{I}$ (with $d_1 = d$) In addition, the elementary step of a layer of *disentangler*s is introduced, so unitaries $U \in U(d_j^2)$ that let respective pairs of nearest neighbours interact, again satisfying $U^\dagger U = \mathbb{I}$,



The rationale behind this approach is that one first disentangles neighbours as much as possible before they are renormalised separately. (i) Again, one has polynomially many degrees of freedom, if $d_{\max} = \max\{d_j\}$, actually $\mathcal{O}(d_{\max}^2 n \log(n))$ many real parameters. (ii) Then contraction is still efficient. This might not be entirely obvious. The key insight is that when computing expectation values $\langle \psi | h_j | \psi \rangle$ for a Hamiltonian term h_j , since all steps are either unitary or isometric, one can remove all tensors outside the *causal cone* of the Hamiltonian term h_j , and the tensors within the causal cone can be sequentially contracted following the temporal order: It is clear by construction that the causal cone will have a finite width



Therefore, the contraction is possible with polynomial effort in n and in d_{\max} . This statement remains true for dimensions $\mathcal{D} > 1$: Hence, MERA constitute an efficiently contractible network in fact in any dimension. It turns out that for $\mathcal{D} > 1$ and cubic lattices, MERA can always be mapped onto a PEPS of finite bond dimension, so the set of MERA states may be viewed as a subset of PEPS [112]. However, MERA can always be exactly contracted efficiently and they exhibit a very special and meaningful structure. (iii) Indeed, MERA can be put into contact with critical systems and *conformal field theories*. In several ways, one can view a MERA as a lattice instance of a conformal field theory, an insight that has been fleshed out in quite some detail [113–115]. First numerical implementations of this idea of a MERA were presented in Refs. [116, 117]; in the meantime, these approaches have also been generalised to higher-dimensional systems [118–121].

4 Fermionic and continuum models

4.1 Fermionic models

4.1.1 Fermionic tensor networks

Models that have *fermionic degrees of freedom* associated with the lattice sites can be mapped onto spin models with \mathbb{C}^2 constituents by virtue of the *Jordan Wigner transformation*. However, only for one-dimensional models this approach is practical, as only then a local fermionic model is mapped again onto a local spin model. For two-dimensional fermionic models on $L^{\mathcal{D}}$ with $\mathcal{D} = 2$, say, in contrast, one encounters highly non-local terms that have a locality region growing as $\Omega(L)$, no matter what specific ordering is chosen, rendering the methods discussed here inappropriate (and the computation of expectation values inefficient). This seems unfortunate, in particular since some of the challenging and at the same time most interesting models such as the two-dimensional fermionic *Hubbard model* with Hamiltonian

$$H = -t \sum_{\langle j,k \rangle, \sigma} (f_{j,\sigma}^\dagger f_{k,\sigma} + \text{h.c.}) + U \sum_j f_{j,\uparrow}^\dagger f_{j,\uparrow} f_{j,\downarrow}^\dagger f_{j,\downarrow} \quad (61)$$

for $t, U \in \mathbb{R}$ are exactly of this kind.

There is a way to overcome this obstacle, however. The prejudice to be overcome is that one should not fix a global order beforehand, but rather update the local order ‘on the fly’, in a way such that all expectation values $\langle O_A \rangle$ are correctly reproduced and give the same values as if one had (inefficiently) mapped the system onto a non-local spin model. This idea has been introduced in Refs. [120–123] and further applied in Refs. [124, 125]. One way of formulating the problem is in terms of *fermionic operator circuits*, more general objects than standard tensor networks that also specify information about the local order of edges, correctly accounting for the signs encountered when *swapping fermionic modes*.

4.1.2 Tensor networks in quantum chemistry

An emergent field of research is the study of systems of *quantum chemistry* with tensor network methods. These are again interacting fermionic models, but this time with a Hamiltonian that is lacking an obvious structure of locality. In second quantisation, Hamiltonians in quantum chemistry can be written as

$$H = \sum_{i,j \in V} T_{i,j} f_i^\dagger f_j + \sum_{i,j,k,l \in V} V_{i,j,k,l} f_i^\dagger f_j^\dagger f_k f_l \quad (62)$$

where V and T are some tensors that do not necessarily reflect geometrically local interactions of spinless fermionic modes. Yet, one can still order the orbitals suitably and consider this now as a one-dimensional quantum system, albeit one with intricate interactions, and run a DMRG algorithm [126, 127]. Also, one can employ tree-tensor network and complete graph approaches [128]. It is an interesting emerging field to make use of tensor network ideas to capture such models of quantum chemistry.

4.2 Continuum models

4.2.1 Quantum fields

We finally very briefly sketch how tensor network methods can be applied to capture continuous systems, as they arise in quantum field theory; obviously, we will not be able to be fair to this subject, but rather provide a glimpse into it. The type of systems that can most naturally be captured in the way described below is constituted by one-dimensional systems of bosons or fermions on line segments of length L , associated with *field operators* $\Psi(x)$ and $\Psi^\dagger(x)$, where $[\Psi(x), \Psi^\dagger(y)] = \delta(x - y)$ for bosons and $\{\Psi(x), \Psi^\dagger(y)\} = \delta(x - y)$ for fermions, for $x, y \in [0, L]$. A common Hamiltonian in this context is the *Lieb-Liniger model*, a non-relativistic model of a continuous bosonic system with a contact interaction, with Hamiltonian

$$H = \int_0^L dx \left(\frac{d\Psi^\dagger(x)}{dx} \frac{d\Psi(x)}{dx} + c\Psi^\dagger(x)\Psi^\dagger(x)\Psi(x)\Psi(x) \right), \quad (63)$$

for some $c > 0$. This model is Bethe-integrable and physically important – it models, for example, the situation of interacting cold bosonic atoms on top of atom chips – so serves as a meaningful benchmark.

4.2.2 Continuous matrix product states

Continuous matrix product states are variational states meaningfully describing such systems. Their state vectors are defined as

$$|\psi\rangle = \text{tr}_2 \left(\mathcal{P} \exp \left(\int_0^L dx (Q(x) \otimes \mathbb{I} + R(x) \otimes \Psi^\dagger(x)) \right) \right) |\emptyset\rangle. \quad (64)$$

where $|\emptyset\rangle$ denotes the vacuum, $\{R(x) \in \mathbb{C}^{D \times D} : x \in [0, L]\}$ and $\{Q(x) \in \mathbb{C}^{D \times D} : x \in [0, L]\}$ are position-dependent matrices reflecting a bond dimension D [129–131], and \mathcal{P} denotes path-ordering. How can they be seen to reflect the meaningful continuum limit of MPS? In the translationally invariant setting – giving for simplicity of notation rise to the situation that R and Q are not position dependent – one can think for a given $L > 0$ of $n = L/\varepsilon$ lattice sites with infinite-dimensional local Hilbert spaces of bosonic modes, and consider the limit $n \rightarrow \infty$. For a given bond dimension D and for $R, Q \in \mathbb{C}^{D \times D}$ one can take as the matrices of the family of translationally invariant MPS

$$A_1 = \mathbb{I} + \varepsilon Q, \quad A_2 = \sqrt{\varepsilon} R, \quad A_k = \sqrt{\varepsilon}^k R^k / k!, \quad (65)$$

for $k \geq 1$, and identify $\Psi_j = a_j / \sqrt{\varepsilon}$ for $j = 1, \dots, n$, which indeed yields Eq. (64) in the limit $n \rightarrow \infty$. Again, in order to compute expectation values of polynomials of field operators, one does not have to perform computations in the physical Hilbert space, but merely in the correlation space of the (continuous) MPS. Equilibrium properties of a one-dimensional quantum field are found to relate to non-equilibrium properties of a zero-dimensional quantum field [129–131].

References

- [1] U. Schollwoeck, *Annals of Physics* **326**, 96 (2011)
- [2] J. Eisert, M. Cramer, and M.B. Plenio, *Rev. Mod. Phys.* **82**, 277 (2010)
- [3] F. Verstraete, J.I. Cirac, and V. Murg, *Adv. Phys.* **57**, 143 (2008)
- [4] N. Schuch: (2013). [arXiv:1306.5551](#)
- [5] R. Orus: (2013). [arXiv:1306.2164](#)
- [6] E.H. Lieb, T. Schultz, and D. Mattis, *Ann. Phys.* **16**, 407 (1961)
- [7] E.H. Lieb and D. W. Robinson, *Commun. Math. Phys.* **28**, 251 (1972)
- [8] M.B. Hastings and T. Koma, *Commun. Math. Phys.* **265**, 781 (2006)
- [9] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters, *Phys. Rev. A* **54**, 3824 (1996)
- [10] M.J. Donald, M. Horodecki, and O. Rudolph, *J. Math. Phys.* **43**, 4252 (2002)
- [11] M. Srednicki, *Phys. Rev. Lett.* **71**, 66 (1993)
- [12] M.B. Hastings, *J. Stat. Mech.*, P08024 (2007)
- [13] I. Arad, A. Kitaev, Z. Landau, and U. Vazirani: (2013). [ArXiv:1301.1162](#)
- [14] M.B. Plenio, J. Eisert, J. Dreissig, and M. Cramer, *Phys. Rev. Lett.* **94**, 060503 (2005)
- [15] M. Cramer and J. Eisert, *New J. Phys.* **8**, 71 (2006)
- [16] M. Cramer, J. Eisert, and M.B. Plenio, *Phys. Rev. Lett.* **98**, 220603 (2007)
- [17] M. Hein, J. Eisert, and H.J. Briegel, *Phys. Rev. A* **69**, 062311 (2004)
- [18] F. Verstraete, M.M. Wolf, D. Perez-Garcia, and J.I. Cirac, *Phys. Rev. Lett.* **96**, 220601 (2006)
- [19] K.V. Acoleyen, M. Marien, and F. Verstraete: (2013). [ArXiv:1304.5931](#)
- [20] S. Michalakis: (2012). [arXiv:1206.6900](#)
- [21] C. Holzhey, F. Larsen, and F. Wilczek, *Nucl. Phys. B* **424**, 443 (1994)
- [22] P. Calabrese and J. Cardy, *J. Stat. Mech.*, P06002 (2004)
- [23] A.R. Its, B.-Q. Jin, and V.E. Korepin, *J. Phys. A* **38**, 2975 (2005)
- [24] F. Franchini, A. R. Its, B.-Q. Jin, and V.E. Korepin, *J. Phys. A* **40**, 8467 (2007)

- [25] J. Eisert and M. Cramer, Phys. Rev. A **72**, 042112 (2005)
- [26] G. Vidal, J.I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003)
- [27] P. Calabrese and J. Cardy, Int. J. Quant. Inf. **4**, 429 (2006)
- [28] M.M. Wolf, Phys. Rev. Lett. **96**, 010404 (2006)
- [29] D. Gioev and I. Klich, Phys. Rev. Lett. **96**, 100503 (2006)
- [30] M.M. Wolf, F. Verstraete, M.B. Hastings, and J.I. Cirac, Phys. Rev. Lett. **100**, 070502 (2008)
- [31] H. Bernigau, M. Kastoryano, and J. Eisert: (2013). arXiv:1301.5646
- [32] J. Eisert, PhD thesis, and University of Potsdam (2001)
- [33] G. Vidal and R.F. Werner, Phys. Rev. A **65**, 032314 (2002)
- [34] M.B. Plenio, Phys. Rev. Lett. **95**, 090503 (2005)
- [35] K.M.R. Audenaert, J. Eisert, M.B. Plenio, and R.F. Werner, Phys. Rev. A **66**, 042327 (2002)
- [36] P. Calabrese, J. Cardy, and E. Tonni, J. Stat. Mech., P02008 (2013)
- [37] C. Castelnovo: (2013). arXiv:1306.4990
- [38] I. Peschel, M. Kaulke, and O. Legeza, Ann. Phys. (Leipzig) **8**, 153 (1999)
- [39] H. Li and F.D.M. Haldane, Phys. Rev. Lett. **101**, 010504 (2008)
- [40] V. Alba, M. Haque, and A.M. Laeuchli, J. Stat. Mech., P08011 (2012)
- [41] D. Poilblanc, N. Schuch, D. Perez-Garcia, and J.I. Cirac, Phys. Rev. B **86**, 014404 (2012)
- [42] N. Laflorencie and D. Poilblanc, Lecture Notes in Physics **645**, 227 (2004)
- [43] D. Poulin, A. Qarry, R.D. Somma, and F. Verstraete, Phys. Rev. Lett. **106**, 170501 (2011)
- [44] S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
- [45] U. Schollwoeck, Rev. Mod. Phys. **77**, 259 (2005)
- [46] S. Oestlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995)
- [47] D. Porras, F. Verstraete, and J.I. Cirac, Phys. Rev. Lett. **93**, 227205 (2004)
- [48] M. Fannes, B. Nachtergaele, and R.F. Werner, Lett. Math. Phys. **25**, 249 (1992)
- [49] A. Kluemper, A. Schadschneider, and J. Zittartz, Europhys. Lett. **24**, 293 (1993)

- [50] W. Hackbusch: *Tensor spaces and numerical tensor calculus* (Springer, 2012)
- [51] D. Perez-Garcia, F. Verstraete, M.M. Wolf, and J.I. Cirac,
Quant. Inf. Comp. **5-6**, 401 (2006)
- [52] G. Vidal, Phys. Rev. Lett. **91**, 147902 (2003)
- [53] R. Huebener, A. Mari, and J. Eisert, Phys. Rev. Lett. **110**, 040401 (2013)
- [54] N. Schuch and J.I. Cirac, Phys. Rev. A **82**, 012314 (2010)
- [55] N. Schuch, J.I. Cirac, and F. Verstraete, Phys. Rev. Lett. **100**, 250501 (2008)
- [56] J. Eisert, Phys. Rev. Lett. **97**, 260501 (2006)
- [57] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004)
- [58] A.J. Daley, C. Kollath, U. Schollwoeck, and G. Vidal, J. Stat. Mech., P04005 (2004)
- [59] S.R. White and A. Feiguin, Phys. Rev. Lett. **93**, 076401 (2004)
- [60] J. Haegeman, J.I. Cirac, T.J. Osborne, I. Pizorn, H. Verschelde, and F. Verstraete,
Phys. Rev. Lett. **107**, 070601 (2011)
- [61] I.P. McCulloch: (2008). arXiv:0804.2509
- [62] G.M. Crosswhite, A.C. Doherty, and G. Vidal, Phys. Rev. B **78**, 035116 (2008)
- [63] R. Orus and G. Vidal, Phys. Rev. B **78**, 155117 (2008)
- [64] F. Verstraete, J.J. Garcia-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004)
- [65] M. Zwolak and G. Vidal, Phys. Rev. Lett. **93**, 207205 (2004)
- [66] M. Znidaric, T. Prosen, and I. Pizorn, Phys. Rev. A **78**, 022103 (2008)
- [67] T. Barthel, New J. Phys. **15**, 073010 (2013)
- [68] C. Kollath, A. Laeuchli, and E. Altman, Phys. Rev. Lett. **98**, 180601 (2007)
- [69] M. Cramer, A. Flesch, I.P. McCulloch, U. Schollwoeck, and J. Eisert,
Phys. Rev. Lett. **101**, 063001 (2008)
- [70] S. Trotzky, Y.-A. Chen, A. Flesch, I.P. McCulloch, U. Schollwoeck, J. Eisert, and
I. Bloch, Nature Physics **8**, 325 (2012)
- [71] T. Fukuhara, A. Kantian, M. Endres, M. Cheneau, P. Schauss, S. Hild, D. Bellem,
U. Schollwoeck, T. Giamarchi, C. Gross, I. Bloch, and S. Kuhr,
Nature Physics **9**, 235 (2013)

- [72] J. Eisert and T.J. Osborne, Phys. Rev. Lett **97**, 150404 (2006)
- [73] N. Schuch, M.M. Wolf, K.G.H. Vollbrecht, and J.I. Cirac, New J. Phys. **10**, 033032 (2008)
- [74] M.C. Banuls, M.B. Hastings, F. Verstraete, and J.I. Cirac, Phys. Rev. Lett. **102**, 240603 (2009)
- [75] N. Schuch, D. Perez-Garcia, and J.I. Cirac, Phys. Rev. B **84**, 165139 (2011)
- [76] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B **83**, 035107 (2011)
- [77] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010)
- [78] M. Sanz, M.M. Wolf, D. Perez-Garcia, and J. I. Cirac, Phys. Rev. A **79**, 042308 (2009)
- [79] A. Cadenas, M. Sanz, M.M. Wolf, J.I. Cirac, and D. Perez-Garcia, Phys. Rev. B **87**, 035114 (2013)
- [80] M.M. Wolf, G. Ortiz, F. Verstraete, and J.I. Cirac, Phys. Rev. Lett. **97**, 110403 (2006)
- [81] I. Affleck, T. Kennedy, E. Lieb, and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987)
- [82] F.D.M. Haldane, Phys. Rev. Lett. **50**, 1153 (1983)
- [83] C.K. Majumdar and D.K. Ghosh, J. Math. Phys. **10**, 1388 (1969)
- [84] D. Perez-Garcia, M.M. Wolf, M. Sanz, F. Verstraete, and J.I. Cirac, Phys. Rev. Lett. **100**, 167202 (2008)
- [85] J. Haegeman, M. Marien, T.J. Osborne, and F. Verstraete: (2012). arXiv:1210.7710
- [86] M. Jarzyna and R. Demkowicz-Dobrzanski, Phys. Rev. Lett. **110**, 240405 (2013)
- [87] R. Raussendorf and H.J. Briegel, Phys. Rev. Lett. **86**, 5188 (2001)
- [88] D. Gross and J. Eisert, Phys. Rev. Lett. **98**, 220503 (2007)
- [89] R. Huebener, C. Kruszynska, L. Hartmann, W. Duer, F. Verstraete, J. Eisert, and M.B. Plenio, Phys. Rev. A **79**, 022317 (2009)
- [90] D. Gross, J. Eisert, N. Schuch, and D. Perez-Garcia, Phys. Rev. A **76**, 052315 (2007)
- [91] T.-C. Wei, I. Affleck, and R. Raussendorf, Phys. Rev. A **86**, 032328 (2012)
- [92] D. Gross and J. Eisert, Phys. Rev. A **82**, 040303 (2010)
- [93] F. Verstraete, M.A. Martin-Delgado, and J.I. Cirac, Phys. Rev. Lett. **92**, 087201 (2004)

- [94] M. Cramer, M.B. Plenio, S.T. Flammia, D. Gross, S.D. Bartlett, R. Somma, O. Landon-Cardinal, Y.-K. Liu, and D. Poulin, *Nat. Commun.* **1**, 149 (2010)
- [95] F. Verstraete and J.I. Cirac, (2004). *cond-mat/0407066*
- [96] T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **67**, 3066 (1998)
- [97] G. Sierra and M.A. Martin-Delgado: *cond-mat/9811170*
- [98] H. Niggemann, A. Kluemper, and J. Zittartz, *Z. Phys. B* **104**, 103 (1997)
- [99] N. Schuch, M.M. Wolf, F. Verstraete, and J.I. Cirac, *Phys. Rev. Lett.* **98**, 140506 (2007)
- [100] Z.-C. Gu, M. Levin, and X.-G. Wen, *Phys. Rev. A* **78**, 205116 (2008)
- [101] H.C. Jiang, Z. Y. Weng, and T. Ziang, *Phys. Rev. Lett.* **101**, 090603 (2008)
- [102] J.I. Cirac, S. Michalakis, D. Perez-Garcia, and N. Schuch: (2013). *arXiv:1306.4003*
- [103] J. Jordan, R. Orus, G. Vidal, F. Verstraete, and J.I. Cirac, *Phys. Rev. Lett.* **101**, 250602 (2008)
- [104] R. Orus, J.I. Latorre, J. Eisert, and M. Cramer, *Phys. Rev. A (R)* **73**, 060303 (2006)
- [105] S. Yan, D.A. Huse, and S.R. White, *Science* **332**, 1173 (2011)
- [106] S. Depenbrock, I.P. McCulloch, and U. Schollwoeck, *Phys. Rev. Lett.* **109**, 067201 (2012)
- [107] D. Poilblanc, N. Schuch, D. Perez-Garcia, and J.I. Cirac, *Phys. Rev. B* **86**, 014404 (2012)
- [108] Y. Shi, L. Duan, and G. Vidal, *Phys. Rev. A* **74**, 022320 (2006)
- [109] P. Silvi, V. Giovannetti, S. Montangero, M. Rizzi, J. I. Cirac, and R. Fazio, *Phys. Rev. A* **81**, 062335 (2010)
- [110] N. de Beaudrap, M. Ohliger, T.J. Osborne, and J. Eisert, *Phys. Rev. Lett.* **105**, 060504 (2010)
- [111] G. Vidal, *Phys. Rev. Lett.* **99**, 220405 (2007)
- [112] T. Barthel, M. Kliesch, and J. Eisert, *Phys. Rev. Lett.* **105**, 010502 (2010)
- [113] V. Giovannetti, S. Montangero, M. Rizzi, and R. Fazio, *Phys. Rev. A* **79**, 052314 (2009)
- [114] R.N.C. Pfeifer, G. Evenbly, and G. Vidal, *Phys. Rev. A (R)* **79**, 040301 (2009)
- [115] B. Swingle: (2009). *arXiv:0905.1317*
- [116] C. M. Dawson, J. Eisert, and T.J. Osborne, *Phys. Rev. Lett.* **100**, 130501 (2008)

- [117] M. Rizzi, S. Montangero, and G. Vidal, *Phys. Rev. A* **77**, 052328 (2008)
- [118] G. Evenbly and G. Vidal, *Phys. Rev. B* **79**, 144108 (2009)
- [119] G. Evenbly and G. Vidal, *Phys. Rev. Lett.* **102**, 180406 (2009)
- [120] P. Corboz and G. Vidal, *Phys. Rev. B* **80**, 165129 (2009)
- [121] C. Pineda, T. Barthel, and J. Eisert, *Phys. Rev. A* **81**, 050303 (2010)
- [122] P. Corboz, G. Evenbly, F. Verstraete, and G. Vidal, *Phys. Rev. A (R)* **81**, 010303 (2010)
- [123] T. Barthel, C. Pineda, and J. Eisert, *Phys. Rev. A (R)* **80**, 042333 (2009)
- [124] Z.-C. Gu, F. Verstraete, and X.-G. Wen, (2010). [arXiv:1004.2563](#)
- [125] I. Pizorn and F. Verstraete, *Phys. Rev. B* **81**, 245110 (2010)
- [126] G. Barcza, O. Legeza, K. H. Marti, and M. Reiher, *Phys. Rev. A* **83**, 012508 (2011)
- [127] K. Boguslawski, P. Tecmer, O. Legeza, and M. Reiher, *J. Phys. Chem. Lett.* **3**, 3129 (2012)
- [128] K.H. Marti, B. Bauer, M. Reiher, M. Troyer, and F. Verstraete, *New J. Phys.* **12**, 103008 (2010)
- [129] F. Verstraete and J.I. Cirac, *Phys. Rev. Lett.* **104**, 190405 (2010)
- [130] T.J. Osborne, J. Eisert, and F. Verstraete, *Phys. Rev. Lett.* **105**, 260401 (2010)
- [131] J. Haegeman, J.I. Cirac, T.J. Osborne, H. Verschelde, and F. Verstraete, *Phys. Rev. Lett.* **105**, 251601 (2010)

Index

A

acceptance probability, 6.5
action
 pair, 14.4
 primitive approximation, 14.3
adiabatic approximation, 12.3
Affleck-Kennedy-Lieb-Tasaki (AKLT) model,
 16.4, 17.24
algorithm
 heatbath, 6.5
 loop, 7.16
 Metropolis, 6.2, 14.4
 Metropolis-Hastings, 6.4
 Swendson-Wang, 7.17
 Wolff, 6.6, 7.17
 worm, 7.21, 14.7
anticommutator, 2.13
antiferromagnetism, 3.15, 3.21, 10.4, 10.5,
 14.12
 itinerant, 10.5
 Slater, 10.5
 strong coupling, 10.5
 weak coupling, 10.5
anyons, 2.4
area law, 17.7

B

back-propagation, 15.18
balance
 detailed, 6.2
 global, 6.2
bath orbitals, 4.18
BCS limit, 13.12
BCS state, 2.23
BCS theory, 11.4, 12.18, 13.2
Born effective charge (BEC), 11.13
Born-Oppenheimer approximation, 12.3
Bose condensation, 14.7

C

Car-Parrinello method, 1.15
CDMFT, 10.18
charge density wave, 9.6, 9.14
charge susceptibility, 12.9
checkerboard decomposition, 7.5

clustering of correlations, 17.5
Coloumb matrix elements, 15.4
complex auxiliary-fields, 15.18
condensate fraction, 14.13
configuration interaction (CI), 2.9
conserving approximations, 4.2
constrained path Monte Carlo (CPMC),
 15.16
continuous matrix product states, 17.35
continuous time, 7.8
continuous-time QMC, 5.14
cooling rate, 8.30
correlation times, 6.7
Coulomb exchange, 3.6
Coulomb pseudopotential, 13.10
Coulson's challenge, 2.7
coupling approach, 6.7
critical point, 6.6
crystallization of Ge/Sb/Te, 1.18
cumulative order parameter, 10.27
Curie temperature, 5.28
cusp condition, 14.4

D

DCA, 10.19
density matrix, 2.6, 14.5
 single particle, 14.13
 thermal, 14.2
density of states, 11.7
density-functional perturbation theory,
 12.9
density-functional theory (DFT), 1.2,
 12.5
density-matrix renormalization group (DMRG),
 8.13, 16.2, 17.21
 dynamical, 16.27
detailed balance, 6.2
diagonal update, 7.15
directed loops, 7.21
dispersion (van der Waals) forces, 1.15,
 1.16
dispersion kink, 12.25
doped ionic insulators, 11.7
dynamic polarizability, 11.8

dynamical matrix, 12.5

Dyson equation, 4.6

E

effective magnetic moment, 3.24

Ehrenfest theorem, 1.5

eigenvalue decomposition (EVD), 16.5

electron density, 1.5

electron self-energy, 12.22

electron-phonon coupling, 11.6, 12.12

electron-phonon interaction, 9.7, 9.13

electron-phonon vertex, 12.13

electronic linewidth, 12.22

electronic self-energy, 4.2

Eliashberg equations

imaginary axis, 13.8

real axis, 13.11

Eliashberg function, 12.18, 13.11

Eliashberg theory, 11.6, 12.18, 13.5

multi-band, 13.21

emergent properties, 10.4

emergent view, 1.2

entangled state, 2.21, 16.3

entanglement, 16.3

entanglement entropy, 16.6, 17.6

entanglement negativity, 17.9

entanglement spectrum, 17.9

entropy

entanglement, 16.6, 17.6

von Neumann, 16.6

exact diagonalization, 8.7, 8.9–8.11, 8.21

exchange cycles, 14.6

exchange frequency, 14.10

exchange hole, 1.6, 2.11

exchange interaction, 5.11

exchange-correlation energy, 1.9, 1.16

exchange-correlation hole, 1.10, 2.8

excited states, 15.24

exotic pairing, 11.5

F

faster-than-the-clock approach, 6.5

Fermi liquid, 14.10

Fermionic models, 17.34

ferromagnet, 14.12

ferromagnetism, 3.12

field operator, 2.14

fluctuation-dissipation theorem, 3.38

fluctuation-exchange approximation (FLEX),

10.8

Fock Space, 2.14

force bias, 15.20

frustration, 10.6

functional renormalization group, 10.12

G

gap equation, 11.4, 13.2, 13.8

generating functional, 4.13

giant spin Hamiltonian, 8.7

Gibbs state, 17.9

grand canonical potential, 4.2

Great Solid State Physics Dream Machine, 1.12

Green function, 4.5, 5.6

H

Hamiltonian

giant spin approximation, 8.7

Heisenberg, 8.6

magnetic anisotropy, 8.6

of magnetic molecules, 8.6

rotational band, 8.23

Zeeman interaction, 8.6

harmonic approximation, 12.5

Hartree approximation, 1.3

Hartree-Fock approximation, 1.3, 1.15, 2.11, 3.25

heatbath algorithm, 6.5

Heisenberg model, 3.6, 5.2, 8.6, 8.20, 14.12, 17.25

helium

liquid ^4He , 14.7

solid ^3He , 14.12

solid ^4He , 14.13

high-temperature superconductivity, 9.2, 9.11–9.13

Hubbard model, 3.7, 5.2, 10.3, 17.34

Hubbard-Stratonovich (HS) transformation, 15.8

hybrid formalism in AFQMC, 15.22

I

imaginary time Green function, 3.33, 4.4

importance sampling, 15.16

improved estimators, 7.20

impurity solver, 10.19
indistinguishability, 2.3
inelastic neutron scattering, 9.3, 9.9
instanton, 14.11
isomorphism, 14.13
isothermal magnetic entropy change, 8.31

K

Kandel-Domany formulation, 7.17
Kohn-Luttinger theorem, 10.12, 11.5
Kohn-Sham theory, 1.7
Kondo effect, 3.28
Kramers-Kronig relations, 3.37

L

Lanczos method, 8.11
 finite-temperature, 8.18
lattice dynamics, 12.4
Legendre transform, 4.17
Lehmann representation, 4.5
linear response, 12.8
 function, 3.38
 theory, 3.36
local updates, 7.7
local-density approximation (LDA), 1.9,
 1.14, 1.16, 1.18
locally optimal solution, 7.23
loop algorithm, 7.16
Luttinger-Ward functional, 4.7

M

magnetic cooling, 8.27
magnetic entropy change, 8.31
magnetic ion, 3.2
magnetic molecules, 8.2
magnetic resonant mode, 9.2, 9.10, 9.11
magnetism
 molecular, 8.2
magnetization
 jumps, 8.26
 tunneling, 8.3, 8.7
magnetocalorics, 8.27
Markov chain, 6.2
Marshall-Peierls sign rule, 8.21
matrix-product operators, 17.22
matrix-product state (MPS), 16.2, 17.13
Matsubara Green function, 3.33
Matsubara sums, 3.34

mean-field approximation, 16.3
Mermin-Wagner-Hohenberg theorem, 10.6
metal-insulator transition, 4.21, 14.12
Metropolis algorithm, 6.2
Metropolis-Hastings algorithm, 6.4
MgB₂, 11.9
Migdal's theorem, 13.7
modified Gram-Schmidt, 15.11
molecular dynamics, 14.4
molecular magnetism, 8.2
momentum distribution, 14.7
Mott transition, 10.6
 crossover, 10.6
Mott-Ioffe-Regel limit, 10.10
multi-scale entanglement renormalisa-
 tion, 17.33
multiband superconductor, 12.20
multiplet theory, 4.26

N

Nambu formalism, 13.5
no sign problem half-filling repulsive Hub-
 bard model, 15.13

O

observables, 8.18
 heat capacity, 8.18
 magnetization, 8.18
 susceptibility, 8.18
odd-membered spin rings, 8.21
operator loops, 7.21
organic crystals, 11.23
oxide heterostructures, 9.14

P

Padé method, 13.13
pair-correlation function, 2.8
paramagnetism of isolated ions, 3.17
particle-hole symmetry, 4.22
partition function, 5.16, 8.18
path integral, 5.14, 14.3
 fermion, 14.8
Pauli paramagnetism, 3.10
Pauli principle, 2.2
periodization, 10.18
permutation operator, 14.5
phase problem, 15.22
phase-change materials, 1.18

phaseless AFQMC, 15.18
 phonon linewidth, 12.16
 phonon self-energy, 12.15
 phonon-mediated pairing interaction, 12.18
 point group C_N , 8.9
 poor-man scaling, 3.29
 Potthoff functional, 10.17
 product basis, 8.7
 product state, 16.3
 projected entangled pair state (PEPS),
 17.15, 17.29
 pseudogap, 9.3, 9.13

Q

quantum critical point, 9.6, 9.7
 quantum fields, 17.35
 quantum lattice model, 17.3
 quantum Monte Carlo methods, 1.17
 quantum mutual information, 17.9
 quantum state tomography, 17.29
 quantum-to-classical mapping, 7.7
 quasiparticle renormalization, 12.22
 quenched dynamics, 17.23

R

random-phase approximation (RPA), 3.16,
 10.8
 rare-earth elements, 1.2
 real materials, 1.3
 reference system, 4.18
 reorthogonalization, 15.16
 resonant elastic x-ray scattering (REXS),
 9.3, 9.6
 resonant inelastic x-ray scattering (RIXS),
 9.3, 9.10, 9.12, 9.13
 resonating valence bond (RVB), 2.22
 restricted path integral identity, 14.8
 rotational band, 8.23

S

screened interactions, 5.11
 second quantization, 2.12
 self-consistent renormalized theory (SCR),
 10.8
 shift operator, 8.9
 sign problem, 5.14, 7.23, 14.8, 15.13
 single-particle basis, 15.6
 single-particle Green's function, 15.7

singular value, 16.5
 singular value decomposition (SVD), 16.5
 skeleton diagrams, 4.10
 Slater determinant, 1.3, 2.8, 15.6
 Slater exchange potential, 1.6
 slave particles, 10.14
 spin fluctuations, 9.2, 9.8
 spin models, 3.21
 spin rings, 8.20
 spin susceptibility, 3.41, 5.27
 spin wave spectrum, 5.21
 spin-fermion model, 9.11, 9.13
 spin-statistics connection, 2.3
 split operator approximation, 7.3
 spontaneously broken symmetry, 4.25
 stochastic series expansion, 7.10
 Stoner instabilities, 3.12
 Stoner interatomic exchange, 5.2
 stripe order, 9.6, 9.14
 strongly-correlated systems, 1.2, 1.17
 sub-Kelvin cooling, 8.27
 superconductivity, 10.9, 11.3, 12.18, 13.2
 Eliashberg, 10.9
 strongly correlated, 10.9
 weakly correlated, 10.9
 superfluid, 14.6, 14.13
 fraction, 14.6
 supersolid, 14.13
 susceptibility, 3.38, 11.7
 Swendsen-Wang algorithm, 7.17
 symmetry, 8.7, 8.8, 8.11
 cyclic groups, 8.9
 translational symmetry, 8.9
 symmetry factor of a diagram, 4.11

T

tensor network, 17.11
 thermal de Broglie wavelength, 14.7
 thermodynamic observables, 8.18
 Thomas-Fermi theory, 1.2, 1.4
 time-dependent variational principle, 17.23
 time-evolving block decimation method,
 17.23
 toric code, 17.32
 torque, 5.20
 transfer operator, 17.18
 transition metal

compounds, 4.26
dichalcogenides, 11.17
nitrochlorides, 11.13
pnictide halides, 11.15
Trotter
 error, 7.8
 time-step, 7.6
Trotter decomposition, 7.3, 14.3, 15.8,
 16.16
tunneling frequency, 14.11
tunnelling inversion, 13.15
two-particle self-consistent theory (TPSC),
 10.8

V

vacuum state, 2.13
variational cluster approximation (VCA),
 4.2
virial estimator, 14.5

W

Wannier functions, 5.5
Wigner crystal, 14.12
winding number, 14.6, 14.13
Wolff algorithm, 6.6, 7.17
world lines, 7.6
worm algorithm, 7.21

X

$X\alpha$ approximation, 1.9

Z

Zeeman interaction, 3.4, 8.6

1. **The LDA+DMFT approach to strongly correlated materials**

Lecture Notes of the Autumn School 2011 Hands-on LDA+DMFT

edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2011), 420 pages

ISBN: 978-3-89336-734-4

2. **Correlated Electrons: From Models to Materials**

Lecture Notes of the Autumn School on Correlated Electrons 2012

edited by E. Pavarini, E. Koch, F. Anders, M. Jarrell (2012), 450 pages

ISBN: 978-3-89336-796-2

3. **Emergent Phenomena in Correlated Matter**

Lecture Notes of the Autumn School on Correlated Electrons 2013

edited by E. Pavarini, E. Koch, U. Schollwöck (2013), 520 pages

ISBN: 978-3-89336-884-6